



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :C12N 15/12, C07K 14/435, 16/18, C12N
5/12, A61K 38/17, C12N 15/63

A1

(11) International Publication Number:

WO 98/16639

(43) International Publication Date:

23 April 1998 (23.04.98)

(21) International Application Number: PCT/US97/17374

(22) International Filing Date: 26 September 1997 (26.09.97)

(30) Priority Data:

60/027,337

11 October 1996 (11.10.96)

US

(71) Applicants (for all designated States except US): SUGEN, INC.
[US/US]; 351 Galveston Drive, Redwood City, CA 94063
(US). NEW YORK UNIVERSITY MEDICAL CENTER
[US/US]; 550 1st Avenue, New York, NY 10016 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): LEV, Sima [IL/US]; 8
Locksley Avenue, San Francisco, CA 94122 (US). PLOW-
MAN, Gregory, D. [US/US]; 4 Honeysuckle Lane, San Car-
los, CA 94070 (US). SCHLESSINGER, Joseph [IL/US]; 37
Washington Square West, New York, NY 10011 (US).

(74) Agents: WARBURG, Richard, J. et al.; Lyon & Lyon LLP,
Suite 4700, 633 West Fifth Street, Los Angeles, CA
90071-2066 (US).

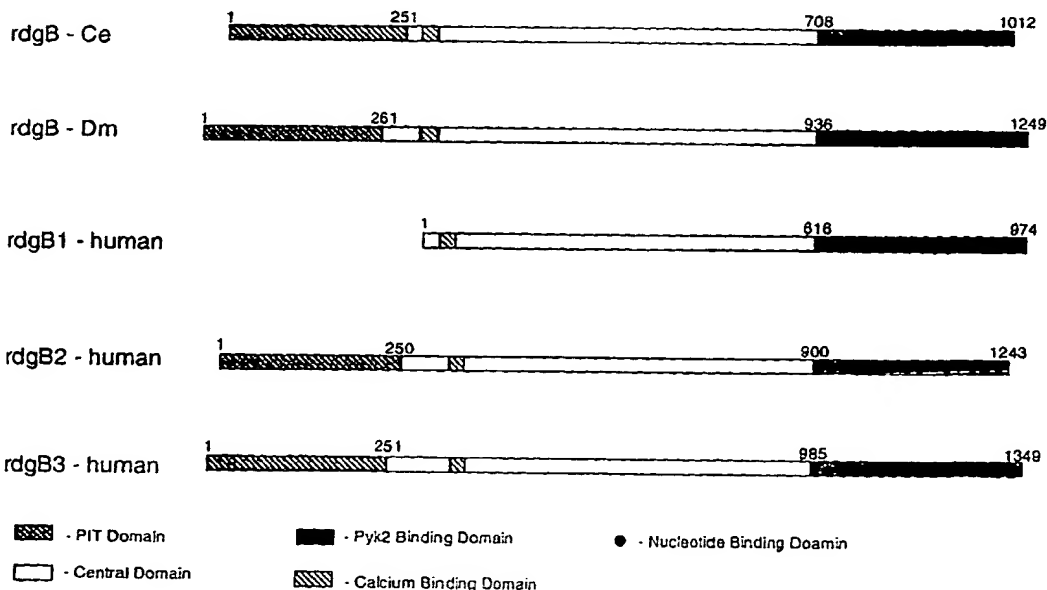
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR,
BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE,
GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO,
NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH,
KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE,
CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,
ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the
claims and to be republished in the event of the receipt of
amendments.

(54) Title: RDGB-PROTEINS



(57) Abstract

The present invention features a method for treatment of an organism having a disease or condition characterized by an abnormality in a signal transduction pathway, wherein the signal transduction pathway includes a rdgB protein. The invention also features methods for diagnosing such diseases and for screening for agents that will be useful in treating such diseases. The invention also features purified and/or isolated nucleic acid encoding a rdgB protein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DESCRIPTION

RDGB-PROTEINS

Introduction

The present invention relates generally to newly identified rdgB proteins and related products and methods.

Background of the Invention

5 The following discussion of the background of the invention and references cited therein are not admitted to be prior art to the invention.

Cellular signal transduction is a fundamental mechanism whereby external stimuli that regulate diverse cellular processes are relayed to the interior of cells. 10 One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of tyrosine residues on proteins. The phosphorylation state of a protein is modified through the reciprocal actions of 15 tyrosine phosphatases (TPs) and tyrosine kinases (TKs), including receptor tyrosine kinases and non-receptor tyrosine kinases.

A tyrosine protein kinase named PYK2, is described in U.S. patent application serial No. 08/460,626, filed June 20 2, 1995, which is a continuation-in-part application of U.S. patent application Serial No. 08/357,642, filed December 15, 1994, both of which are hereby incorporated herein by reference in their entirety including any drawings. PYK2 contains an N-terminal domain, a catalytic 25 domain, two proline-rich regions, potential Src homology 2 (SH2) binding regions, and a region homologous to the focal adhesion targeting domain.

A type of protein found in *Drosophila*, called *Drosophila* retinal degeneration B protein (rdgB) is 30 described in Vihtelic et al., *J. of Cell Biology* 122, :1013-1022, 1993. The sequence described in this reference, however, contained a false stop codon

sequencing error and thus the authors were not aware that the *Drosophila* rdgB contains a PYK-2 binding domain. In addition, this sequence was incorrectly identified as a member of the 6-transmembrane domain family of proteins.

5 These rdgB proteins function in many sensory and neuronal cells of the fly and are directly associated with sight in the fly.

The sequence of a genomic clone of a portion of *C. elegans* has been placed on a computer database, and
10 (although unappreciated), this sequence contains an rdgB sequence with introns. Thus, the GENE BANK database contains raw data of the nucleotide sequence of a series of genomic clones of *c. Elegans*. Using portions of the human rdgb sequence, the present invention identifies an
15 open reading frame that has been to this point unrecognized. An rdgB was thus found segregated into 14 exons in two separate cosmids C54C6 (assc. #Z77131) and MO1F1 (assc. #Z46381).

Summary of the Invention

20 The present invention relates to rdgB polypeptides, nucleic acids encoding such polypeptides, cells, tissues and animals containing such polypeptides, antibodies to such polypeptides, assays utilizing such polypeptides, and methods relating to all of the foregoing. Such rdgB
25 polypeptides are involved in various signal transduction pathways and thus the present invention provides several agents and methods useful for diagnosing, treating, and preventing various diseases or conditions associated with abnormalities in these pathways.

30 The present invention is based in part upon the identification and isolation of a series of novel non-receptor tyrosine kinase binding molecules, termed hrdgB1, hrdgB2, and hrdgB3. The full length nucleic acid sequences encoding these proteins are set forth
35 respectively in SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3. The full length amino acid sequences are set forth

respectively in SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6. RDGBs are generally comprised of 3 structural domains. The N-terminal PIT domains described herein have approximately 45% amino acid identity to human PPI1 and
5 PPI2. The PIT domains of RDGB2 and RDGB3 (RDGB1 lacks a PIT domain) have approximately 72% identity with each other and approximately 62-65% identity with the drosophila and C elegans rdgB's. The full length amino acid sequence for c. Elagans is set forth in SEQ ID NO:7
10 and the full length Drosophila nucleic acid sequence set forth in SEQ ID NO:8, and the full length Drosophila amino acid sequence is set forth in SEQ ID NO:9. The PIT domains of the rdgBs have a conserved putative ATP binding motif similar to that seen in protein kinases.

15 The second central domain is present in all human rdgbs described herein and has no sequence homology to any other known domain. The three human rgdbs share 43-47% identity over the 600 to 675 amino acid stretch and show 25-35% identity to the invertebrate rdgB's. This large
20 domain contains three subdomains with much higher identity (66-88% in the human rdgbs and 35-75% with the invertebrate rdbgs.) This high level of conservation, especially across such a diverse set of species, suggests an important functional role for these stretches. The N-
25 terminal portion of the central domain is a conserved acidic region of 10 to 15 amino acids comprised almost exclusively of glutamatic and aspartate residues that may function as a calcium binding motif.

The third rdgB domain is particularly unique to these
30 proteins and consists of the C-terminal 343 to 384 residues of the proteins. There is 60-63% identity amongst the human rdgbs and 40-60% with the invertebrate rdgB's. The comparison with the drosophila rdgb is based on the unique knowledge of this domain and its functional
35 significance as described herein. The published sequence contained a framseshift mutation such that the protein was previously thought to terminate less than halfway through

this domain. By comparison with the human sequences, the present invention provides a sequence that extends beyond the end of the drosophila sequence to include amino acids 1054-1249.

5 Within the PYK2 binding domain is a distinct motif with primary sequence homology to the nucleotide binding region of the ras-related GTP-binding proteins. All members of this family (ras, rho, rac, rab, ran) contain a sequence characterized by the conserved hydrophobic-
10 hydrophobic-G-X-K-X-D-hydrophobic amino acid sequence. The G-X-K motif in the rdgBs is at aa 614 (rdgb1), aa898 (rdgb2), aa 983 (rdgb3) and aa 987 (dm). Based on analysis of the three dimensional structure (by X-ray crystallography) of this region from ras and ran, this
15 motif grasps the nucleotide ring of GDP/GTP as part of the molecular "on-off" switch in these proteins. The rdgbs however lack the upstream p-llop or A-box present in these small G-proteins.

RdgB proteins are involved in key signal transduction
20 pathways related to neurotransmitter signaling. This is based in part on the recognition of existence and significance of domains found in rdgB proteins (see Figure 1). For example, the experiments described herein demonstrate that rdgB proteins contain a PYK2 binding
25 domain. PYK2 is believed to be responsible for regulating neurotransmitter signaling. The rdgB proteins also contain a PIT domain, which in Drosophila is involved in PI transfer. PI transfer in humans is involved in the recycling of synaptic vesicles. Thus, in view of the
30 roles of the PYK2 binding domain and the PIT domain, rdgB proteins may be useful in the treatment of conditions of nervous system by enhancing or inhibiting such signaling.

Thus, in a first aspect the invention features an isolated, purified, enriched or recombinant nucleic acid
35 encoding a rdgB polypeptide. Preferably such nucleic acid encodes a mammalian rdgB polypeptide, more preferably it encodes a human rdgB polypeptide.

By "isolated" in reference to nucleic acid is meant a polymer of 2 (preferably 21, more preferably 39, most preferably 75) or more nucleotides conjugated to each other, including DNA or RNA that is isolated from a natural source or that is synthesized. The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but does indicate that it is the predominate sequence present (at least 10 - 20% more than any other nucleotide sequence) and is essentially free (about 90 - 95% pure at least) of non-nucleotide material naturally associated with it. Therefore, the term does not encompass an isolated chromosome encoding one or more rdgB polypeptides.

By the use of the term "enriched" in reference to nucleic acid is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction (2 - 5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased in a useful manner and preferably separate from a sequence library. The term significant here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other nucleic

acids of about at least 2 fold, more preferably at least 5 to 10 fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, e.g., in terms of mg/ml). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6 -fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more

preferably four or five orders of magnitude is expressly contemplated.

By "rdgB polypeptide" is meant 9 or more contiguous amino acids set forth in the full length amino acid sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6. The rdgB polypeptides can be encoded by full-length nucleic acid sequences or any portion of a full-length nucleic acid sequence, so long as a functional activity of the polypeptide is retained. Preferred functional activities include the ability to bind to the N-terminal portion of PYK2. For example, the present invention encompasses deletion mutants isolated domains, and complementary sequences capable of hybridizing to full length rdgB protein under stringent hybridization conditions.

In preferred embodiments, isolated nucleic acid comprises, consists essentially of, or consists of a nucleic acid sequence set forth in the full length nucleic acid sequence SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 or at least 27, 30, 45, 60 or 90 contiguous nucleotides thereof and the rdgB polypeptide comprises, consists essentially of, or consists of at least 9, 10, 15, 20, 30, 50, 100, 200, or 300 contiguous amino acids of a rdgB polypeptide.

By "comprising" it is meant including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase

"consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

Compositions and probes of the present invention may contain human nucleic acids encoding a rdgB polypeptide but are substantially free of nucleic acid not encoding rdgB polypeptide. The human nucleic acid encoding a rdgB polypeptide is at least 18 contiguous bases of the nucleotide sequence set forth in SEQ. ID NO. 1, SEQ. ID NO. 2, or SEQ. ID NO. 3 and will selectively hybridize to human genomic DNA encoding a rdgB polypeptide, or is complementary to such a sequence. The nucleic acid may be isolated from a natural source by cDNA cloning or subtractive hybridization; the natural source may be blood, semen, and tissue of various organisms including eukaryotes, mammals, birds, fish, plants, gorillas, rhesus monkeys, chimpanzees and humans; and the nucleic acid may be synthesized by the triester method or by using an automated DNA synthesizer. In yet other preferred embodiments the nucleic acid is a conserved or unique region, for example those useful for the design of hybridization probes to facilitate identification and cloning of additional polypeptides, the design of PCR probes to facilitate cloning of additional polypeptides, and obtaining antibodies to polypeptide regions.

By "conserved nucleic acid regions", are meant regions present on two or more nucleic acids encoding a rdgB polypeptide, to which a particular nucleic acid sequence can hybridize to under lower stringency conditions. Examples of lower stringency conditions suitable for screening for nucleic acid encoding rdgB polypeptides are provided in Abe, et al. J. Biol. Chem., 19:13361 (1992) (hereby incorporated by reference herein in its entirety, including any drawings). Preferably, conserved regions differ by no more than 7 out of 20 nucleotides.

By "unique nucleic acid region" is meant a sequence present in a full length nucleic acid coding for a rdgB polypeptide that is not present in a sequence coding for any other naturally occurring polypeptide. Such regions
5 preferably comprise 12 or 20 contiguous nucleotides present in the full length nucleic acid encoding a rdgB polypeptide.

The invention also features a nucleic acid probe for the detection of a rdgB polypeptide or nucleic acid
10 encoding a rdgB polypeptide in a sample. The nucleic acid probe contains nucleic acid that will hybridize to at least one sequence set forth in SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In preferred embodiments the nucleic acid probe
15 hybridizes to nucleic acid encoding at least 12, 27, 30, 35, 40, 50, 100, 200, or 300 contiguous amino acids of the full-length sequence set forth in SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6. Various low or high stringency hybridization conditions may be used depending upon the
20 specificity and selectivity desired.

By "high stringency hybridization conditions" is meant those hybridizing conditions that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at
25 50°C; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C; or
30 (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M Sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS. Under stringent hybridization conditions only highly
35 complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of

nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides.

Methods for using the probes include detecting the presence or amount of rdgB RNA in a sample by contacting
5 the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to rdgB RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a rdgB polypeptide may be used in the
10 identification of the sequence of the nucleic acid detected (for example see, Nelson et al., in Nonisotopic DNA Probe Techniques, p. 275 Academic Press, San Diego (Kricka, ed., 1992) hereby incorporated by reference herein in its entirety, including any drawings). Kits for
15 performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant
20 nucleic acid may contain a sequence set forth in SEQ ID NO:1 and a vector or a promoter effective to initiate transcription in a host cell. The recombinant nucleic acid can alternatively contain a transcriptional initiation region functional in a cell, a sequence
25 complimentary to an RNA sequence encoding a rdgB polypeptide and a transcriptional termination region functional in a cell.

In another aspect the invention features an isolated, enriched or purified rdgB polypeptide.

30 By "isolated" in reference to a polypeptide is meant a polymer of 2 (preferably 7, more preferably 13, most preferably 25) or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are synthesized. The isolated
35 polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a

naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is the predominate sequence present (at least 10 - 20% more than any other sequence) and is essentially free (about 90 - 95% pure at least) of non-amino acid material naturally associated with it.

10 By the use of the term "enriched" in reference to a polypeptide is meant that the specific amino acid sequence constitutes a significantly higher fraction (2 - 5 fold) of the total of amino acids present in the cells or solution of interest than in normal or diseased cells or
15 in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acids present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two.
20 However, it should be noted that enriched does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term significant here is used to indicate that the level of increase is useful
25 to the person making such an increase, and generally means an increase relative to other amino acids of about at least 2 fold, more preferably at least 5 to 10 fold or even more. The term also does not imply that there is no amino acid from other sources. The other source amino
30 acid may, for example, comprise amino acid encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which man has intervened to elevate the proportion of the desired amino acid.

35 It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require

absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, e.g., in terms of mg/ml). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of contamination at a functionally significant level, for example 90%, 95%, or 99% pure.

In preferred embodiments rdgB polypeptides contain at least 9, 10, 15, 20, or 30 contiguous amino acids of the full-length sequence set forth in SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6.

In yet another aspect the invention features a purified antibody (e.g., a monoclonal or polyclonal antibody) having specific binding affinity to a rdgB polypeptide. The antibody contains a sequence of amino acids that is able to specifically bind to a rdgB polypeptide.

By "specific binding affinity" is meant that the antibody will bind to a hrgdB polypeptide at a certain detectable amount but will not bind other polypeptides to the same extent, under identical conditions. The present invention also encompasses antibodies that can distinguish hrgdB1 from hrgdB2 or hrgdB3 or can otherwise distinguish between the various rdgBs.

Antibodies having specific binding affinity to a rdgB polypeptide may be used in methods for detecting the presence and/or amount of a rdgB polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the rdgB polypeptide. Diagnostic kits for performing such methods may be constructed to include a first container means containing the antibody and a second container means

having a conjugate of a binding partner of the antibody and a label.

In another aspect the invention features a hybridoma which produces an antibody having specific binding
5 affinity to a rdgB polypeptide.

By "hybridoma" is meant an immortalized cell line which is capable of secreting an antibody, for example a rdgB antibody.

In preferred embodiments the rdgB antibody comprises
10 a sequence of amino acids that is able to specifically bind a rdgB polypeptide.

Another aspect of the invention features a method of detecting the presence or amount of a compound capable of binding to a rdgB polypeptide. The method involves
15 incubating the compound with a rdgB polypeptide and detecting the presence or amount of the compound bound to the rdgB polypeptide.

In preferred embodiments, the compound inhibits an activity of rdgB. The present invention also features
20 compounds capable of binding and inhibiting rdgB polypeptide that are identified by methods described above.

In another aspect the invention features a method of screening potential agents useful for treatment of a
25 disease or condition characterized by an abnormality in a signal transduction pathway that contains an interaction between a rdgB polypeptide and a natural binding partner (NBP). The method involves assaying potential agents for those able to promote or disrupt the interaction as an
30 indication of a useful agent.

By "screening" is meant investigating an organism for the presence or absence of a property. The process may include measuring or detecting various properties, including the level of signal transduction and the level
35 of interaction between a rdgB polypeptide and a NBP.

By "disease or condition" is meant a state in an organism, e.g., a human, which is recognized as abnormal

by members of the medical community. The disease or condition may be characterized by an abnormality in one or more signal transduction pathways in a cell, preferably a cell listed in table 1, wherein one of the components of the signal transduction pathway is either a rdgB polypeptide or a NBP.

Specific diseases or disorders which might be treated or prevented, based upon the affected cells include: myasthenia gravis; neuroblastoma; disorders caused by neuronal toxins such as cholera toxin, pertussis toxin, or snake venom; acute megakaryocytic myelosis; thrombocytopenia; those of the central nervous system such as seizures, stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell damage such as in cardiac arrest or neonatal distress, epilepsy, neurodegenerative diseases such as Alzheimer's disease, Huntington's disease and Parkinson's disease, dementia, muscle tension, depression, anxiety, panic disorder, obsessive-compulsive disorder, post-traumatic stress disorder, schizophrenia, neuroleptic malignant syndrome, and Tourette's syndrome. Conditions that may be treated by rdgB inhibitors include epilepsy, schizophrenia, extreme hyperactivity in children, chronic pain, and acute pain. Examples of conditions that may be treated by PYK2-rdgB pathway enhancers (for example a phosphatase inhibitor) include stroke, Alzheimer's, Parkinson's, other neurodegenerative diseases and migraine.

Preferred disorders include epilepsy, stroke, schizophrenia, and Parkinson's disorder as there is an established relationship between these disorders and the function of potassium channels. See, McLean et al., Epilepsia 35:S5-S9 1994; Ricard-Mousnier et al., Neurophysiologie Clinique 23:395-421, 1993; Crit. Rev. Neurobiol 7:187-203, 1994; Simon and Lin, Biophys. J. 64:A100, 1993; Birnstiel et al., Synapse (NY) 11:191-196, 1992; Coleman et al., Brain Res. 575:138-142 1992; Popolip et al., Br. J. Pharmacol 104:907-913, 1991; Murphy et al.,

Exp. Brain Res. 84:355-358, 1991; Rutecki et al., Epilepsia 32:1-2, 1991; Fisher and Coyle (ed), Frontiers of Clinical Neuroscience, Vol. 11 "Neurotransmitters and Epilepsy"; Meeting, Woods Hole MA, USA IX+260P. John Wiley and Sons, Inc. NY, NY; Treherne and Ashford, Neuroscience 40:523-532, 1991; Gehlert, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 18:1093-1102, 1994; Baudy, Expert Opin Ther. Pat. 1994 4/4:343-378; Porter and Rogawski, Epilepsia 33:S1-S6, 1992; Murphy, J. Physiol. 453:167-183, 1992; Cromakalim, Drugs Future 17/3:237-239, 1992; Carmeliet, Eur. Heart J. 12:30-37, 1991; Olpe et al., Experientia 47/3:254-257, 1991; Andrade et al., Science 234/4781:1261-1265, 1986; Forster, J. Neurosci. Methods 13/3-4:199-212, 1985.

15 In preferred embodiments, the methods described herein involve identifying a patient in need of treatment. Those skilled in the art will recognize that various techniques may be used to identify such patients. For example, cellular potassium levels may be measured or the
20 individuals genes may be examined for a defect.

By "abnormality" is meant an a level which is statistically different from the level observed in organisms not suffering from such a disease or condition and may be characterized as either an excess amount, intensity
25 or duration of signal or a deficient amount, intensity or duration of signal. The abnormality in signal transduction may be realized as an abnormality in cell function, viability or differentiation state. The present invention is based in part on the determination that such
30 abnormality in a pathway can be alleviated by action at the PYK2-rdgB interaction site in the pathway. An abnormal interaction level may also either be greater or less than the normal level and may impair the normal performance or function of the organism. Thus, it is also
35 possible to screen for agents that will be useful for treating a disease or condition, characterized by an abnormality in the signal transduction pathway, by testing

compounds for their ability to affect the interaction between a rdgB polypeptide and PYK2, since the complex formed by such interaction is part of the signal transduction pathway. However, the disease or condition may be characterized by an abnormality in the signal transduction pathway even if the level of interaction between the rdgB polypeptide and NBP is normal.

By "interact" is meant any physical association between polypeptides, whether covalent or non-covalent. This linkage can include many chemical mechanisms, for instance covalent binding, affinity binding, intercalation, coordinate binding and complexation. Examples of non-covalent bonds include electrostatic bonds, hydrogen bonds, and Van der Waals bonds. Furthermore, the interactions between polypeptides may either be direct or indirect. Thus, the association between two given polypeptides may be achieved with an intermediary agent, or several such agents, that connects the two proteins of interest (e.g., a rdgB polypeptide and PYK2). Another example of an indirect interaction is the independent production, stimulation, or inhibition of both a rdgB polypeptide and PYK2 by a regulatory agent. Depending upon the type of interaction present, various methods may be used to measure the level of interaction. For example, the strengths of covalent bonds are often measured in terms of the energy required to break a certain number of bonds (i.e., kcal/mol). Non-covalent interactions are often described as above, and also in terms of the distance between the interacting molecules. Indirect interactions may be described in a number of ways, including the number of intermediary agents involved, or the degree of control exercised over the rdgB polypeptide relative to the control exercised over PYK2 or another NBP.

By "disrupt" is meant that the interaction between the rdgB polypeptide and PYK2 or a NBP is reduced either by preventing expression of the rdgB polypeptide, or by

preventing expression of PYK2 or NBP, or by specifically preventing interaction of the naturally synthesized proteins or by interfering with the interaction of the proteins.

5 By "promote" is meant that the interaction between a rdgB polypeptide and PYK2 or NBP is increased either by increasing expression of a rdgB polypeptide, or by increasing expression of PYK2 or a NBP, or by decreasing the dephosphorylating activity of the corresponding
10 regulatory PTP (or other phosphatase acting on other phosphorylated signaling components) by promoting interaction of the rdgB polypeptide and PYK2 or NBP or by prolonging the duration of the interaction. Covalent binding can be promoted either by direct condensation of
15 existing side chains or by the incorporation of external bridging molecules. Many bivalent or polyvalent linking agents are useful in coupling polypeptides, such as an antibody, to other molecules. For example, representative coupling agents can include organic compounds such as
20 thioesters, carbodiimides, succinimide esters, diisocyanates, glutaraldehydes, diazobenzenes and hexamethylene diamines. This listing is not intended to be exhaustive of the various classes of coupling agents known in the art but, rather, is exemplary of the more
25 common coupling agents. (See Killen and Lindstrom 1984, J. Immunol. 133:1335-2549; Jansen, F.K., et al., 1982, Immunological Rev. 62:185-216; and Vitetta et al., supra).

By "NBP" is meant a natural binding partner of a rdgB polypeptide that naturally associates with a rdgB
30 polypeptide. The structure (primary, secondary, or tertiary) of the particular natural binding partner will influence the particular type of interaction between the rdgB polypeptide and the natural binding partner. For example, if the natural binding partner comprises a
35 sequence of amino acids complementary to the rdgB polypeptide, covalent bonding may be a possible interaction. Similarly, other structural characteristics

may allow for other corresponding interactions. The interaction is not limited to particular residues and specifically may involve phosphotyrosine, phosphoserine, or phosphothreonine residues. A broad range of sequences
5 may be capable of interacting with rdgB polypeptides. One example of a natural binding partner may be pyk2, which is described above. Using techniques well known in the art, one may identify several natural binding partners for rdgB polypeptides such as by utilizing a two-hybrid screen.

10 By "signal transduction pathway" is meant the sequence of events that involves the transmission of a message from an extracellular protein to the cytoplasm through a cell membrane. The signal ultimately will cause the cell to perform a particular function, for example, to
15 uncontrollably proliferate and therefore cause cancer. Various mechanisms for the signal transduction pathway (Fry et al., Protein Science, 2:1785-1797, 1993) provide possible methods for measuring the amount or intensity of a given signal. Depending upon the particular disease
20 associated with the abnormality in a signal transduction pathway, various symptoms may be detected. Those skilled in the art recognize those symptoms that are associated with the various other diseases described herein. Furthermore, since some adapter molecules recruit
25 secondary signal transducer proteins towards the membrane, one measure of signal transduction is the concentration and localization of various proteins and complexes. In addition, conformational changes that are involved in the transmission of a signal may be observed using circular
30 dichroism and fluorescence studies.

In another aspect the invention features a method of diagnosis of an organism for a disease or condition characterized by an abnormality in a signal transduction pathway that contains an interaction between a rdgB
35 polypeptide and PYK2 or a NBP. The method involves detecting the level of interaction as an indication of said disease or condition.

By "organism" is meant any living creature. The term includes mammals, and specifically humans. Preferred organisms include mice, as the ability to treat or diagnose mice is often predictive of the ability to function in other organisms such as humans.

By "diagnosis" is meant any method of identifying a symptom normally associated with a given disease or condition. Thus, an initial diagnosis may be conclusively established as correct by the use of additional confirmatory evidence such as the presence of other symptoms. Current classification of various diseases and conditions is constantly changing as more is learned about the mechanisms causing the diseases or conditions. Thus, the detection of an important symptom, such as the detection of an abnormal level of interaction between rdgB polypeptides and PYK2 or NBPs may form the basis to define and diagnose a newly named disease or condition. For example, conventional cancers are classified according to the presence of a particular set of symptoms. However, a subset of these symptoms may both be associated with an abnormality in a particular signalling pathway, such as the ras²¹ pathway and in the future these diseases may be reclassified as ras²¹ pathway diseases regardless of the particular symptoms observed.

Yet another aspect of the invention features a method for treatment of an organism having a disease or condition characterized by an abnormality in a signal transduction pathway. The signal transduction pathway contains an interaction between a rdgB polypeptide and PYK2 or a NBP and the method involves promoting or disrupting the interaction, including methods that target the rdgB:NBP interaction directly, as well as methods that target other points along the pathway.

By "dominant negative mutant protein" is meant a mutant protein that interferes with the normal signal transduction pathway. The dominant negative mutant protein contains the domain of interest (e.g., an rdgB

polypeptide or PYK2 or a NBP), but has a mutation preventing proper signaling, for example by preventing binding of a second domain from the same protein. One example of a dominant negative protein is described in
5 Millauer et al., Nature February 10, 1994. The agent is preferably a peptide which blocks or promotes interaction of the rdgB polypeptide and PYK2 or another NBP. The peptide may be recombinant, purified, or placed in a pharmaceutically acceptable carrier or diluent.

10 An EC_{50} or IC_{50} of less than or equal to 100 μM is preferable, and even more preferably less than or equal to 50 μM , and most preferably less than or equal to 20 μM . Such lower EC_{50} 's or IC_{50} 's are advantageous since they allow lower concentrations of molecules to be used *in vivo*
15 or *in vitro* for therapy or diagnosis. The discovery of molecules with such low EC_{50} 's and IC_{50} 's enables the design and synthesis of additional molecules having similar potency and effectiveness. In addition, the molecule may have an EC_{50} or IC_{50} less than or equal to 100 μM at one or
20 more, but not all cells chosen from the group consisting of parathyroid cell, bone osteoclast, juxtaglomerular kidney cell, proximal tubule kidney cell, distal tubule kidney cell, cell of the thick ascending limb of Henle's loop and/or collecting duct, central nervous system cell,
25 keratinocyte in the epidermis, parafollicular cell in the thyroid (C-cell), intestinal cell, trophoblast in the placenta, platelet, vascular smooth muscle cell, cardiac atrial cell, gastrin-secreting cell, glucagon-secreting cell, kidney mesangial cell, mammary cell, beta cell,
30 fat/adipose cell, immune cell and GI tract cell.

By "therapeutically effective amount" is meant an amount of a pharmaceutical composition having a therapeutically relevant effect. A therapeutically relevant effect relieves to some extent one or more
35 symptoms of the disease or condition in the patient; or returns to normal either partially or completely one or more physiological or biochemical parameters associated

with or causative of the disease or condition. Generally, a therapeutically effective amount is between about 1 nmole and 1 μ mole of the molecule, depending on its EC_{50} or IC_{50} and on the age and size of the patient, and the
5 disease associated with the patient.

In another aspect, the invention describes a polypeptide comprising a recombinant rdgB polypeptide or a unique fragment thereof. By "unique fragment," is meant an amino acid sequence present in a full-length rdgB
10 polypeptide that is not present in any other naturally occurring polypeptide. Preferably, such a sequence comprises 6 contiguous amino acids present in the full sequence. More preferably, such a sequence comprises 12 contiguous amino acids present in the full sequence. Even
15 more preferably, such a sequence comprises 18 contiguous amino acids present in the full sequence.

By "recombinant rdgB polypeptide" is meant to include a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide
20 either in its location (e.g., present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

25 In another aspect, the invention describes a recombinant cell or tissue containing a purified nucleic acid coding for a rdgB polypeptide. In such cells, the nucleic acid may be under the control of its genomic regulatory elements, or may be under the control of
30 exogenous regulatory elements including an exogenous promoter. By "exogenous" it is meant a promoter that is not normally coupled *in vivo* transcriptionally to the coding sequence for the rdgB polypeptide.

In another aspect, the invention features a rdgB
35 polypeptide binding agent able to bind to a rdgB polypeptide. The binding agent is preferably a purified antibody which recognizes an epitope present on a rdgB

polypeptide. Other binding agents include molecules which bind to the rdgB polypeptide and analogous molecules which bind to a rdgB polypeptide.

By "purified" in reference to an antibody is meant
5 that the antibody is distinct from naturally occurring antibody, such as in a purified form. Preferably, the antibody is provided as a homogeneous preparation by standard techniques. Uses of antibodies to the cloned polypeptide include those to be used as therapeutics, or
10 as diagnostic tools.

In another aspect, the invention provides a nucleic acid molecule comprising a nucleotide sequence that encodes: (a) a polypeptide having an amino acid sequence set forth in SEQ ID NO:4 from amino acid residues 1-616 or
15 616-974; (b) the complement of the nucleotide sequence of (a); (c) a polypeptide having an amino acid sequence set forth in SEQ ID NO:5 from amino acid residues 1-250, 250-900, or 900-1243; (d) the complement of the nucleotide sequence of (c); (e) a polypeptide having an amino acid
20 sequence of SEQ ID NO:6 from amino acid residues 1-251, 251-985, or 985-1349; or (f) the complement of the nucleotide sequence of (e). The utility of such isolated domains in the design of protein inhibitors is well-known to those skilled in the art.

25 The invention also provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO:4; SEQ ID NO:5, or SEQ ID NO:6 except that it lacks at least one, but not more than two, of the
30 domains selected from the group consisting of the PIT, the central domain, the PYK2 binding domain, the calcium binding domain and the nucleotide binding domain. Such deletion mutants are useful in the design of assays for protein inhibitors. The nucleic acid molecules described
35 above may be, for example, cDNA or genomic DNA and may be placed in a recombinant vector or expression vector. In such a vector, the nucleic acid preferably is operatively

associated with the regulatory nucleotide sequence containing transcriptional and translational regulatory information that controls expression of the nucleotide sequence in a host cell.

5 Thus, the invention also provides a genetically engineered host cell containing any of the nucleotide sequences described herein and the nucleic acid preferably is operatively associated with the regulatory nucleotide sequence containing transcriptional and translational
10 regulatory information that controls expression of the nucleotide sequence in a host cell. Such host cells may obviously be either prokaryotic or eukaryotic.

Other features and advantages of the invention will be apparent from the following description of the
15 preferred embodiments thereof, and from the claims.

Brief Description of the Figures

Figure 1 shows the domains of some preferred full length rdgB proteins.

Description of the Preferred Embodiments

20 The present invention relates to rdgB polypeptides, nucleic acids encoding such polypeptides, cells, tissues and animals containing such nucleic acids, antibodies to such polypeptides, assays utilizing such polypeptides, and methods relating to all of the foregoing. Those skilled
25 in the art will recognize that many of the methods described below in relation to rdgB, PYK-2, a NBP, or a complex of rdgB with PYK-2 or a NBP could also be utilized with respect to the other members of this group.

We describe the isolation and characterization of a
30 novel non-receptor tyrosine kinase binding protein, termed rdgB. HrdgB1 is expressed in the brain, spleen, and ovary. HrdgB2 is expressed in many human tissues including brain, heart, thymus, and peripheral blood leukocytes. HrdgB3 is highly expressed in the thymus but
35 is also expressed in the brain, heart, ovary, and testis.

The examples presented for PYK2, supra, reveal a novel mechanism for the coupling, between G-protein coupled receptors and the MAP kinase signaling pathway. These examples also showed that calcium influx induced by
5 membrane depolarization following activation of the nicotinic acetylcholine receptor or other stimuli that cause calcium influx or release from internal stores lead to the activation of PYK2, tyrosine phosphorylation of Shc, recruitment of Grb2/Sos and activation of the MAP
10 kinase signaling pathway. Pyk2 can also link extracellular signals with the JNK/SAP kinase signaling pathway.

RdgB proteins represent a link in the observations disclosed above. RdgB proteins are shown to bind to PYK2
15 with high affinity both in vitro and in vivo. Evidence of this high affinity interaction is visualized in experiments pulling PYK2 out of a cell lysate with glutathione S-transferase fused rdgB proteins. These experiments are described in the Examples section below.
20 In addition the Drosophila homologs of the rdgB proteins contain a phosphatidylinositol transferase domain as well as a Ca²⁺ binding domain. Although the phosphatidyl inositol transferase domain is missing in an alternatively spliced variant, all forms of rdgB proteins contain a Ca²⁺
25 binding domain. Thus the Ca²⁺ binding domain of rdgB proteins are potentially involved in the Ca²⁺ response observed in PYK2 signaling.

The model presented herein may represent the mechanism underlying calcium mediated regulation of gene
30 expression in neuronal cells induced by MMDA receptor or voltage sensitive calcium channels. The expression pattern of PYK2, the external stimuli that activate the kinase together with its role in the control of MAP kinase and JNK signaling pathways suggests a potential role for
35 PYK2 and rdgB proteins in the control of a broad array of processes in the central nervous system including neuronal plasticity, highly localized control of ion channel

function, as well as, localized activation of the MAP kinase and JNK signaling pathways, cell excitability, and synaptic efficacy.

Various other features and aspects of the invention are: Nucleic Acid Encoding A rdgB Polypeptide; A Nucleic Acid Probe for the Detection of RdgB; Probe Based Method And Kit For Detecting RdgB; DNA Constructs Comprising a RdgB Nucleic Acid Molecule and Cells Containing These Constructs; Purified rdgB Polypeptides; RdgB Antibody And Hybridoma; An Antibody Based Method And Kit For Detecting RdgB; Isolation of Compounds Which Interact With RdgB; Compositions; Disruption of Protein Complexes; Antibodies to Complexes; Pharmaceutical Formulations and Modes of Administration; Identification of Agents; Purification and Production of Complexes; Derivatives of Complexes; and Evaluation of Disorders. All of these aspects and features are explained in detail with respect to PYK-2 in PCT publication WO 96/18738, which is incorporated herein by reference in its entirety, including any drawings. Those skilled in the art will readily appreciate that such description can be easily adapted to rgdB as well, and is equally applicable to the present invention.

Examples

The examples below are non-limiting and are merely representative of various aspects and features of the procedures used to identify the full-length nucleic and amino acid sequences of a series of rdgB proteins. Experiments demonstrating rdgB expression, interaction and signalling activities are also provided.

30 Material and Methods

Two hybrid screen

The yeast strain L40 containing the reporter genes HIS3 and β -gal under control of upstream LexA-binding site, was used as a host for the two-hybrid screening. PYK2-N terminal domain (aa 2-245), PYKN- Δ I (aa 2-237),

PYK-NN (aa 2-285) and Fak (aa 2-412) N-terminal domain (aa 2-412) were fused in frame to LexA DNA binding domain. Yeast strain that express the LexA-PYKN fusion protein was transfected with human brain cDNA library (Clontech #HL404AB) fused to GAL4 transcriptional activation domain. Transformants were plated on agar selection medium lacking Uracil (Ura-), Tryptophane (Trp-), Leucine (Leu-) and Histidine (His-). Resulting colonies were isolated and retested for growth on -Ura-Trp-Leu-His plates and for β -galactosidase activity. Plasmid DNA was purified from colonies that were His+, β -gal+ and used for retransformation of yeast strains expressing heterologous baits to determine the specificity of the interaction.

Isolation of rhdgBs cDNAs

hrdgB1: Human brain, Substantia nigra cDNA library (λ gt10, Clontech HL1179a.) was screened with 32-p-labelled probe derived from the yeast prey plasmid encoding GAL10-rhdgB1. Four independent clones were isolated, subcloned and analyzed by sequence. Sequence analysis indicated that the 5' end of the gene is missing from our clones. Therefore human fetal brain cDNA library (λ gt11, clontech HL3003b) was screened with probe derived from the most 5' region of our new cDNA contig. Sequence analysis of six independent clones that were isolated indicated that all of them are belong to the same gene, rhdgB1, but they are missing the 5' end of the gene. A specific-primed cDNA library was constructed in λ ZapII utilizing human fetal brain Poly(A)+ RNA as templet for our cDNA synthcasis (Stratagene Kit). 15 independent clones were isolated and allowed subsequently isolation of the full length cDNA of rhdgB1.

hrdgB2 and rhdgB3: A DNA fragment derived from an EST fragment (T12574) was amplified by PCR from human fetal brain cDNA. The PCR product was subcloned, sequenced and used as a probe for screening a human fetal brain cDNA library (λ gt11, Clontech H15015b). One

positive clone was obtained from this screen. Sequence analysis indicated that it is a partial cDNA clone of a novel gene belongs to the human rdgB family. The cDNA insert of this clone (1.8kb) was used as a probe for
5 rescreening the same cDNA library. Seven independent clones were obtained, subcloned and sequenced. Sequence analysis indicated that all of them belong to the same gene; hrdgB2, but they are different from the original clone that was isolated from the same library. The 3' end
10 of our first clone (1.8kb), was used as a probe to screen a human heart cDNA library (Clontech 7759-1, 7760-1) and allowed subsequent isolation two alternative spliced isoforms of hrdgB3.

Northern blot

15 Human multiple tissues Northern blots (Clontech HL11296) were hybridized under high-stringency conditions using 32P-labelled cDNA fragment of hrdgB1 (EcoRI-Eco47III nuc# 245-511, hrdgB2 (SacI-Eco47III nuc# 1540-2661) and hrdgB3 Bst-X1 nuc# 912-1472 as probe according to the
20 instructions of the manufacture.

Plasmid Constructs-Two-hybrid constructs:

Fusion with LexA DNA-binding domain : PCR was used to amplified different regions of PYK2 and Fak cDNAs as indicated, the amplified DNA fragments were subcloned into
25 pBTM116 in frame to generate a fusion protein with LexA DNA-binding domain.

Fusion with GAL4 activation domain: PCR was used to amplified different regions of hrdgB1, hrdgB2 or hrdgB3 cDNAs as indicated, the amplified DNA fragments were
30 subcloned into pGAD10 (Clontech) in frame to generate a fusion protein with GAL4 activation domain.

Expression vectors

The full length cDNAs of hrdgB1, hrdgB2 and hrdgB3 were subcloned into pCMP1 downstream to CMV promoter. An

HA-epitope tag (YPYDVDPDYAS) SEQ ID NO:10 was fused in frame to their carboxy terminal ends. The PYK2 binding domain of hrdgB2 (residues 911-1243) was subcloned into pCMV-NEO which encode an initiator methionine codon
5 followed by a Myc epitope tag (EQKLISEEDL) SEQ ID NO:1 immediately upstream to the cloning site.

Antibodies

Antibodies against rdgB1 were raised in rabbit immunized either with a synthetic peptide corresponding to
10 amino-acids 965-974 of hrdgB1 (C-Ter Ab), or with a GST-fusion protein containing residues 231-374 (N-Ter Ab). Antibodies against hrdgB2 were raised in rabbit immunized with a synthetic peptide corresponding to amino acids 152-163 of hrdgB2. Antibodies against hrdgB3 were raised in
15 rabbit against MBP-fusion protein containing residues 7-116 of hrdgB3.

Example 1:

Isolation of human rdgB proteins

The yeast two-hybrid system was used to identify
20 proteins that interact with the amino-terminal domain of PYK2. The N-terminal domain of PYK2 was fused to the LexA DNA binding domain and screened a human brain cDNA library. Using a His synthetase gene (HIS3) under the control of LexA operators as a reporter, 124 His⁺ colonies
25 were identified from an initial screen of a million transformants. Of these, 24 were also b-galactosidase positives (gal⁺). Retransformation of these clones into a yeast strain expressing the LexA-PYK2-N fusion protein indicated that only one interacts with the PYK2 N-terminal
30 domain (PYK2-N). The specificity of the interaction was further determined by transformation of this clone into a yeast strain expressing heterologous baits. An interaction was detected in yeast strain expressing either the PYK-N terminal domain, or a shorter version of PYK-N
35 that was missing 48 amino acids from its C-terminal end.

No interaction, however, was detected in strains expressing either the PYK-NN (amino acids 2-285), or the N-terminal domain of Fak, suggesting that this interaction is very specific.

5 The clone that scored for specific interaction with
PYK2-N contained a partial cDNA which allowed subsequent
isolation of a 3.1 kb cDNA with an open reading frame of
975 amino acids. The coding region was flanked by 5' and
3' untranslated regions of 93 and 149bp respectively. The
10 5' untranslated region contains triplet repeats (CGG), a
motif that was identified in many neuropsychiatric
disorders. This region showed homology to the
untranslated region of the human Fragile X mental
retardation FMR-1 gene (66.3% match) using the Smith-
15 Waterman algorithm.

A BLAST search with the full length cDNA sequence
revealed that this protein is related to the drosophila
retinal degeneration B protein (rdgB) and therefore it was
named hrdgB1. The drosophila rdgB protein has an important
20 role in phototransduction pathway. The rdgB mutant was
initially identified by defects in the compound eye, in
that rdgB mutant flies undergo light-enhanced
photoreceptor cell degeneration. The drosophila rdgB
protein contains a phosphatidylinositol transfer domain
25 (PI-TP) in its N-terminal portion, and a calcium binding
site downstream. The protein contains six hydrophobic
regions that were identified as transmembrane domains.
The same hydrophobic regions are conserved in the hrdgB1
protein, however, analysis of rdgB1 sequence, as well as
30 the drosophila homolog, using different algorithms
(PROSITE) indicated that they are not classical
transmembrane domains.

An ESTs data base search with drosophila rdgB
sequence allowed the identification of two additional
35 human genes that belong to the same gene family. A PCR
fragment derived from an EST fragment (T12574) was used as
probe to screen a human brain cDNA library and subsequent

isolation the hrdgB2 gene. The full length cDNA of hrdgB2 (4186 bp) contained an open reading of 1244 amino acids which was flanked by a 5' untranslated region of 174bp and a 3' untranslated region of 280bp. The 257 amino-acids in the Nterminal end of the hrdgB2 protein have 41% similarity to the entire human PtdInsTP (M73704).

The full length cDNA of hrdgB3 was obtained by screening human brain and heart cDNA libraries. An initial clone of 1.8kb was isolated from a human brain library using the PCR product derived from EST fragment (T12574) as a probe. A cDNA fragment derived from our 1.8kb clone was used as a probe to screen a human heart cDNA library and allowed subsequent isolation of hrdgB3 gene. Two isoforms arising from alternative splicing have been identified by cDNA cloning, the longest which encodes a protein of 1349 amino-acids with a predicted molecular weight of 150kDa, and a shorter one which lacks amino-acids 50-378, with a predicted molecular weight of 120kDa. The coding sequence is flanked by a 79bp 5' untranslated region and a 945 bp 3' untranslated region. The N-terminal region of hrdgB3 contains a PI-TP domain that is missing from the alternative spliced isoform. A stretch of glycines and serines was identified within amino acids 612-634 (78% glycine, 22% serine).

Multiple alignment analysis of the novel hrdgB1, hrdgB2 and hrdgB3 revealed high similarity in their primary structure: a PI-TP domain in the amino-terminal region, six conserved hydrophobic regions and very conserved C-terminal region. Unlike the other rdgB family members, hrdgB1 does not contain PtdInsTP domain, this may suggest that our clone represent an alternative spliced isoform.

Example 2:

Tissue distribution of human rdgBs

The levels of hrdgB1, hrdgB2 and hrdgB3 mRNA expression were determined by Northern analysis of various

human tissues. HrdgB1 has a very restricted expression pattern. It is expressed in the brain, spleen and ovary as a message of approximately 7.5kb. By contrast, hrdgB2 is highly expressed in many human tissues as a message of 4.5 kb. Highest levels of expression were detected in the brain, heart, thymus and peripheral blood leukocytes. HrdgB3 is very highly expressed in the thymus, but it is also expressed in the heart, brain, ovary and testis. Two messages were detected for hrdgB3: 7.5kb and 9.5kb messages that may represent the two alternative spliced isoforms that were isolated. The results discussed above indicate the rdgBs gene family members have very different expression patterns, whereas hrdgB1 is very rare, hrdgB2 is abundant and hrdgB3 has a unique pattern of expression.

Example 3:

Mapping the minimal interaction domain of rdgB proteins

To map the PYK2 interaction domain within the hrdgB1 protein, a series of hrdgB1-deletion mutants were constructed and their ability to interact with PYK2-N was tested utilizing the two hybrid system. Our original two hybrid clone containing amino acids 627-975 of hrdgB1 was used as a positive control. Deletion mutants were constructed, and among all these mutants, only hrdgB1- Δ IV, containing amino acids 627-936, interacts with PYK2-N terminal domain. The interaction of this domain with PYK2 was further confirmed by an in vitro binding experiment, showing binding of PYK2 to immobilized GST-fusion protein containing the same portion of hrdgB1. No binding was detected, however, to the GST-protein alone or between hrdgB1- Δ IV mutant and the focal adhesion kinase.

Since hrdgB1 shares high homology with hrdgB2 and hrdgB3 in their C-terminal domains, whether the corresponding regions of these two proteins interact with PYK2 was examined. For this purpose amino acids 911-1244 and 996-1350 of hrdgB2 and hrdgB3 respectively, were fused in frame to the activator domain of Gal-4, and their

ability to interact with PYK2-N was tested by the two hybrid system. The results indicate that hrdgB2 can strongly bind to PYK2 N-terminal domain, whereas the interaction of rdgB3 with PYK2 is quite weak.

5 To further confirm this interaction in vivo, hrdgB2-HA or hrdgB3-HA were coexpressed either with PYK2 or with Fak in COS cells. Following cell lysis, hrdgB proteins were immunoprecipitated by anti-HA antibodies and the presence of PYK2 or Fak in the immunocomplexes was
10 determined by immunoblotting with antibodies against PYK2 or Fak respectively. The results indicate that both hrdgB2 and hrdgB3 interact with PYK2 in vivo. No interaction, however, was detected with the related kinase Fak, suggesting that hrdgBs proteins interact strongly and
15 specifically with PYK2.

To explore whether the 'PYK2 binding domain' of hrdgBs is sufficient to confer association of those two proteins in vivo, a myc-tagged version of the hrdgB2 'PYK2-binding domain' was coexpressed either with PYK2 or
20 with Fak in COS cells, and their interaction was analyzed. The results showed that this domain can interact with PYK2 in vivo and therefore represent a separate domain in this family of proteins.

Example 4:

25 In vivo association of rdgB1 and PYK2

To confirm the interaction of hrdgB1 and PYK2 in vivo an hemagglutinin-tagged rdgB1 and PYK2 were coexpressed in
293 cells. The results indicate that hrdgB1 strongly associates with PYK2. Association of hrdgB1 with the
30 related kinase Fak could not be detected under the same experimental conditions, suggesting a strong and specific interaction of hrdgB1 and PYK2.

To further characterize the interaction between hrdgB and PYK2, an adult rat brain was used as a source of these
35 two proteins. When hrdgB1 was immunoprecipitated from a rat brain homogenate, utilizing specific antibodies

against hrdgB1, PYK2 could be detected in the immunocomplex. However, the stoichiometry of PYK2/rdgB1 interaction was not as high as shown in transfected cells. These results indicate that PYK2 and rdgB1 interact in vivo under physiological condition, and this interaction may have an important regulatory function in the brain.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Lev, Sima

5 (ii) TITLE OF INVENTION: RDGB PROTEINS AND RELATED PRODUCTS AND METHODS

(iii) NUMBER OF SEQUENCES: 11

(iv) CORRESPONDENCE ADDRESS:

10 (A) ADDRESSEE: Lyon & Lyon
(B) STREET: 633 West Fifth Street
Suite 4700
(C) CITY: Los Angeles
(D) STATE: California
(E) COUNTRY: U.S.A.
15 (F) ZIP: 90071-2066

(v) COMPUTER READABLE FORM:

20 (A) MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: IBM P.C. DOS 5.0
(D) SOFTWARE: FastSeq

(vi) CURRENT APPLICATION DATA:

25 (A) APPLICATION NUMBER: To Be Assigned
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

30 (A) NAME: Warburg, Richard J.
(B) REGISTRATION NUMBER: 32,327
(C) REFERENCE/DOCKET NUMBER: 222/105

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (213) 489-1600
 (B) TELEFAX: (213) 955-0440
 (C) TELEX: 67-3510

5 (2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3109 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

	GCGGCGGCGG	CTGCGGTGGC	GGCAGCGAGG	CGAGCGGGGG	GGGGGCGCGG	GCGCGGCGGCT	60
	CGGAGTCCGT	TCGGGGGCCG	AGGCGGTTCG	GGCCGGGGCC	GGGAAGCGCG	AGGAGCGCGC	120
	GTAGCCGCCG	GAGCCCCCGC	CCCGGGACAT	GGCCAAAGCG	GGCCGTGCAG	GTGGTCTCTC	180
15	CCCCGGGCGG	GGTGCCCCCT	GGCACCTTCG	AAATGTCTCT	AGTGACTCTG	TGGAGAGCTC	240
	AGATGATGAA	TTCTTTGATG	CCAGAGAGGA	GATGGCTGAA	GGGAAGAATG	CCATCCTCAT	300
	TGGGATGAGC	CAGTGGAAC	CCAATGACCT	CGTGGAGCAG	ATCGAGACCA	TGGGGAAACT	360
	GGACGAGCAT	CAAGGAGAAG	GGACCGCGCC	GTGCACATCC	AGCATCCTCC	AGGAGAAGCA	420
	GCGAGAACTG	TACCGGGTTT	CCTTGAGAAG	ACAGAGGTTC	CCAGCCCAGG	GAAGCATCGA	480
20	GATCCACGAA	GACAGCGAGG	AAGGCTGCCC	GCAGCGCTCC	TGCAAGACAC	ATGTCTCTCT	540
	GCTGGTCTTG	CATGGGGGAA	ACATCCTGGA	CACGGGTGCC	GGGGACCCGT	CCTGCAAGGC	600
	AGCCGACATC	CACACCTTCA	GCTCCGTGCT	GGAGAAGGTC	ACACGAGCCC	ATTTCCCTGC	660
	TGCCCTGGGC	CACATCCTCA	TCAAGTTCGT	CCCCTGTCTT	GCCATCTGCT	CTGAGGCTTT	720
	CTCGCTTGTC	TCTCACCTGA	ACCCCTACAG	CCACGATGAG	GGCTGCCTCA	GCAGCAGCCA	780
25	GGACCACGTC	CCTCTGGCCG	CCCTTCCCTT	GTGGCCATC	TCCTCCCCGC	AGTACCAGGA	840
	TGCTGTGCGC	ACCGTCATCG	AGCGAGCCAA	CCAGGTCATC	AGAGAGTTCC	TGAAGTCTCT	900
	TGATGGGATT	GGCTTCAGTG	GGCAGGTGTG	TCTCATCGGG	GACTGTGTGG	GGGGCCTCCT	960
	GGCCTTCGAT	GCCATCTGCT	ACAGTGCGGG	GCCCTCAGGG	GACAGCCCTG	CCAGCAGCAG	1020
	CCGGAAGGGG	AGCATCAGCA	GCACCCAGGA	CACCCAGTTC	GCGGTGGAGG	AAGATTGCAG	1080
30	CCTGGCCAGC	AGCAAGCGTC	TCAGCAAAAG	CAACATTGAC	ATCTCCAGTG	GGTTGGAGGA	1140
	TGAGGAGCCC	AAGAGGCCGT	TGCCGCGGAA	ACAGAGCGAC	TCCTCCACCT	ATGACTGCGA	1200
	GGCCATCACC	CAGCACCATG	CCTTCTCTCT	AAGCATCCAC	TCCAGCGTGC	TAAAGGATGA	1260
	GTCTGAGACC	CCGGCGGCTG	GGGGGCCGCA	GCTCCCTGAG	GTCAGCCTGG	GCCGCTTTGA	1320
	CTTCGATGTC	TCCGACTTCT	TCCTCTTCGG	CTCGCCACTG	GGCCTGGTCC	TGGCCATGCG	1380
35	GAGGACGGTG	CTGCCTGGGC	TGGACGGCTT	CCAGGTGCGT	CCTGCCTGCA	GCCAGGTCTA	1440
	CAGCTTCTTC	CATTGCGCAG	ACCCCTCTGC	CTCACGGCTC	GAGCCACTGC	TGGAGCCCAA	1500
	GTTCCACCTG	GTGCCGCTG	TCAGCGTGCC	TCGCTACCAG	AGGTTCCAC	TGGGCGATGG	1560
	GCAGTCCCTC	CTCCTCGCTG	ATGCCCTACA	CACCCACAGC	CCCCTCTTCC	TGGAGGGCAG	1620
	CTCCCGGGAC	AGCCCGCCAC	TTCTGGATGC	CCCTGCCTCG	CCCCCTCAGG	CCTCGAGGTT	1680
40	CCAGCGCCCA	GGACGGAGGA	TGAGCGAGGG	GAGCTCCAC	AGCGAGAGCT	CGGAGTCCTC	1740
	GGACAGCATG	GCACCCGTGG	GTGCCTCCCG	CATCACAGCC	AAGTGGTGGG	GAAGCAAGAG	1800
	GATCAGATAT	GCCCTGTACT	GCCCTGATGT	CCTCACGGCC	TTCCCACCCG	TGGCCTTGCC	1860
	CCACCTCTTC	CACGCCAGTT	ACTGGGAGTC	CACAGACGTG	GTGGCCTTCA	TCCTGAGACA	1920
	GGTAATGCGC	TATGAGAGCG	TGAACATCAA	GGAAAGCGCC	CGCCTGGACC	CTGCAGCACT	1980
45	GAGTCTTGCC	AACCCCCGGG	AGAAGTGGCT	TCGTAAGCGG	ACTCAGGTCA	AGCTGAGGAA	2040
	TGTCACGGCT	AATCACCGGG	CCAATGATGT	GATTGCTGCT	GAAGATGGCC	CCCAGGTCTT	2100
	GGTGGGGCGG	TTCATGTACG	GGCCCTCGA	CATGGTGGCT	CTGACTGGAG	AGAAGGTGGA	2160
	CATCCTAGTA	ATGGCAGAGC	CATCCTCAGG	CCGTGGGTA	CACCTGGACA	CAGAGATCAC	2220
	CAACAGCAGT	GGTCGCATCA	CATACAATGT	GCCGCGGCC	CGGCGCCTGG	GGGTGGGTGT	2280
50	CTATCCTGTG	AAGATGGTGC	TCAGGGGCGA	CCAGACCTGT	GCCATGAGCT	ACCTCACGGT	2340
	GTTGCCCGAG	GGCATGGAGT	GTGTAGTGTT	CAGCATTTGAT	GGGTCTCTCG	CGGCCAGCGT	2400
	GTCTATCATG	GGAAGCGACC	CCAAGGTCCG	GCCGGGTGCA	GTGGATGTTG	TCCGGCACTG	2460
	GCAGGACTTG	GGTACATGA	TCCTTTACAT	CACGGGACGG	CCGGACATGC	AGAAGCAGCG	2520
	GGTGGTGTCT	TGGCTGTCCC	AGCACAACCT	CCCACAGGGC	ATGATCTTCT	TCTCCGACGG	2580
55	GCTGGTGCAT	GACCCGCTGC	GGCAGAAAGG	CATCTTCTCT	CGCAACCTCA	TGCAGGAGTG	2640
	CTTCATCAAA	ATCAGTGCAG	CCTATGGCTC	CACGAAGGAC	ATCTCTGTCT	ACAGCGTGCT	2700
	GGGCCTGCCT	GCCTCCCGAG	TCTTCATTGT	GGGCCGGCCC	ACCAAGAAGT	ACCAAACCCA	2760
	GTGCCAGTTC	CTGAGCGAGG	GCTACGCGCG	ACACTTGGCC	GTGCTGGAGG	CCAGCCACCG	2820
	CTCACGCCCC	AAGAAGAACA	ACTCGCGCAT	GATCTGCGCG	AAGGGCAGCT	TCCGGCTGCA	2880
60	CGCGCAGCCA	GAGTTCCTGC	GGAAGCGCAA	CCACCTGCGC	AGAACCATGT	CAGTGCAGCA	2940
	GCCCCAGCCC	CCCGCCGCCA	ACCCCAAGCT	CGAGCGGGCC	CAGAGCCAGC	CCGAGTCGGA	3000
	CAAAGACCAC	GAGCGGCCCG	TGCCGCGCGC	CAGTGGGCGG	CGTGGGCCCC	CCAAGTTCGA	3060
	GTCGGTGCCT	TGAGGGGTGG	GCTGTGCTCA	GAGCAGGGAG	CGGGGGCCG		3109

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 4190 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CCGGCACTGC GCCTCGGGAG GGTCCGGCCA CCGCTGGAAC CCGAGGCCGG GGCTGGGGGC 60
GCTCCGGGCT CCGACCCACG GGCCGGCCGG CCCTGCCCGG GCTGGGTGAG GGGCGCCCGC 120
10 CTCAAGCTAG AGGAGGAGCG GAGGCCCGCG GCGGCCCGCC GAGCGCCTTC AGGATGCTCA 180
TCAAGGAATA CCACATTCTG CTGCCCATGA GCCTGGACGA GTACCAGGTG GCCCAGCTCT 240
ACATGATCCA GAAAAAGAGC CGGGAGGAGT CTAGTGGTGA GGGCAGCGGC GTGGAGATCC 300
TGGCCAACCG GCCCTACACG GATGGGCCCG GGGGCAGCGG GCAATACACA CACAAGGTGT 360
ACCACGTGGG CTCCACATC CCAGGCTGGT TCCGGGCACT GCTGCCCAAG GCTGCCCTGC 420
15 AGGTAGAAGA GGAATCCTGG AATGCCTACC CCTACACCCG AACCCGGTAC ACCTGCCCTT 480
TCGTGGAGAA ATTCTCCATT GAAATTGAGA CATTATTACCT GCCTGATGGG GGGCAGCAGC 540
CAAACGTCTT CAACCTGAGC GGGGCCGAGA GGAGACAGCG CATCCTGGAC ACCATCGACA 600
TCGTGCGGGA TGCAGTGGCC CCAGGCGAGT ACAAGCAGA AGAGGACCCC CGGCTTTATC 660
ACTCGTCAA GACGGGCCGA GGGCCACTGT CTGATGACTG GGCACGGACG CGGCACAGA 720
20 CCGGGCCCCC TATGTGTGCG TATAAGCTGT GCAAGGTGTA GTTCCGCTAC TGGGGCATGC 780
AAGCCAAGAT CGAGCAGTTC ATCCATGATG TAGGTCTGCG TCGGGTGATG CTGCGGGCCC 840
ACCGCCAGGC CTGGTGTCTG CAGGATGAGT GGACAGAGCT GAGCATGGCT GACATCCGGG 900
CACTGGAAGA GGAGACTGCT CGCATGCTGG CCCAGCGCAT GGCCAAGTGC AACACAGGCA 960
GTGAGGGGTC CGAGGCCCGC CCCCCCGGGA AACCGAGCAC CGAGGCCCGG TCTGCGGCCA 1020
25 GCAACACTGG CACCCCCGAT GGGCCCTGAG CCCCCCAGG CCCAGATGCC TCCCCCGATG 1080
CCAGCTTTGG GAAGCAGTGG TCCTCATCCT CCGGTTCTCT CTAATCATCC CAACATGGAG 1140
GGGCTGTGTC TCCCCAGAGC TTGTCTGAGT GGCGCATGCA GAACATTGCC CGAGACTCTG 1200
AGAACAGCTC CGAGGAAGAG TTCTTTGATG CCCACGAAGG CTTCTCGGAC AGTGAGGAGG 1260
TCTTCCCCAA GGAGATGACC AAGTGGAACT CCAATGACTT CATTGATGCC TTTGCCTCCC 1320
30 CAGTGAGGCG AGAGGGAACG CCAGAGCCTG CAGCCGAGGC AGCTAAAGGC ATTGAGGATG 1380
GGGCCCAAGC ACCCAGGGAC TCAGAGGGCC TGGATGGAGC CGGGGAGCTG GGGGCTGAGG 1440
CATGCGCAGT CCACGCCCTC TTCCTTATCC TGCACAGCGG CAACATCCTG GACTCAGGCC 1500
CTGGAGACGC CAACTCCAAG CAGGCGGATG TGACAGCGCT GAGCTCCGCC TTCGAGGCCG 1560
TCACCCGCAT CCACTTCCCT GAGGCCCTTG GCCACGTGGC GCTGCGACTG GTGCCCTGTC 1620
35 CACCATCTG CGCCGCCGCC TATGCCCTTG TCTCCAACCT GAGCCCTTAC AGCCACGATG 1680
GGGACAGCCT GTCTCGCTCC CAAGACCACA TTCCATGGC TGCCCTGCCA TGCTGTGCCA 1740
CCTCATCCTC CCGCTACCAG GGCGCCGTGG CCACCGTCAT TGCCCGCACC AACCAAGCCT 1800
ACTCAGCCTT CCTGCGCTCA CCTGAGGGTG CCGGCTTCTG TGGGCAGGTC GCACTGATTG 1860
40 GAGATGGTGT TGGTGGCATC CTGGGCTTTG ATGCACTCTG CCACAGTGCT AACCGGGGCA 1920
CCGGGAGTCG GGGCAGCAGC CGCGCTGGGA GCATGAACAA TGAGCTGCTC TCTCCGGAGT 1980
TTGGCCAGT GCGGGACCCC CTGGCAGATG GTGTGGAAGG CCTGGGTCGG GGCGAGCCAG 2040
AACCCTCGGC CTTGCCCTCCC CAGCGCATCC CCAGCGACAT GGCCAGTCCT GAGCCCGAGG 2100
GCTCTCAGAA CAGCCTTCAG GCAGCCCCCG CACCCAGCTC CTCCTGGGAG CCGCGGCCGG 2160
45 CAAGCACGGC CTTCTGCCCA CCCGCTGCCA GTTCCGAGGC ACCTGACGGC CCCAGCAGCA 2220
CTGCCCGCCT TGAATTCAG GTCTCTGGT TCTTCTCTT CCGCTCCCCA CTGGGCTGCG 2280
TGCTGGCTCT CCGCAAACT GTGATGCCCG CCTTGAGGGC AGCCAGATG CGCCAGCCT 2340
GTGAACAGAT CTACAACCTC TTCCACGCGG CCGACCCCTG CGCCTCACGC CTCGAGCCCC 2400
TGCTGGCCCC GAAGTTCAG GCCATCGCCC CACTGACCGT GCGCCGCTAC CAGAAGTTCC 2460
50 CCGTGGGAGA TGGCTCATCC CTGCTGCTGG CCGACACTCT GCAGACGCAC TCCAGCCTCT 2520
TTCTGGAGGA GCTGGAGATG CTGGTGCCCT CAACACCCAC CTCTACTAGC GTGCTCTCT 2580
GGAAGGGCAG TGAGTTGGCC ACTGACCCCG CCGCCAGGCC AGCCGCCCCC AGCACCACCA 2640
GTGAGGTGGT TAAGATCCTG GAGCGCTGGT GGGGGACCAA GCGGATCGAC TACTCGCTGT 2700
ACTGCCCGCA GCGGCTCACC GCCTTTCCCA CCGTCAAGCT GCGCCACCTC TTCCAGCCCA 2760
55 GCTACTGGGA GTCCGCCGAC GTGGTGGCGT TCATCTGCGC CCAGGTGATC GAGAAGGAGC 2820
GGCCACAGCT GCGGGAATGC GAGGAGCCGT CCACTTACAG CCGGCCCTTC CCCAGGGAGA 2880
AGTGGCAGCG AAAACGCACG CAGGTCAAGA TCCGGAACGT CACTTCCAAC CACCGGCCGA 2940
CGGACACGGT GGTGTGCGAG GGGCCGCCCG AGGTGCTAAG CCGGCGCTTC ATGTACGGGC 3000
CCCTGACGCT CGTCACGCTC ACTGGAGAGA AGGTGGATGT CTACATCATG ACGCAGCCCG 3060
60 TTCCGGGCAA GTGGATCCAC TTTGGCACCG AAGTCAACCA TAGCTCGGGC CGCCTCACCT 3120
TGCTGAGTTC CCCAGAAGCG CGCTGGGGA TTGGTGTCTA CCGCGTGGCG ATGGTGGTCA 3180
GGGGCGACCA CACCTATGCC GAATGCTGCC TGAATGTGGT GGCCCGCGGC ACGGAGGCTG 3240
TGGTCTTCAG CATCGACGGC TCCTTACCGC CAGCGTCTC CATCATGGGC AGCGACCCCA 3300
AGGTGCGAGC TGGCGCGGTG GACGTGGTCA GGCATCTGGC GGAATCCGGC TACTGATCG 3360
65 TGTATGTCAC AGGCCGGCCG GATATGCAGA AGCACCGCGT GGTGGCATGG CTGTGCGAGC 3420
ACAACCTCCC CCACGGCGTC GTCTCTCTTC GCGACGGCCT CACCCACGAC CCACTACGCC 3480
AGAAGGCAAT GTTCTGTCAG AGCCTGTTGT AGGAGGTAGA ACTGAACATC GTGGCCGGTT 3540
ATGGGTCTCC CAAAGATGTG GCTGTATACG CCGCGCTGGG GCTGTCCCCG AGCCAGACCT 3600

37

	ACATCGTGGG	CCGTGCCGTG	CGGAAGCTAC	AGGCGCAGTG	CCAGTTCCTG	TCAGACGGCT	3660
	ATGTGGCCCA	CCTGGGCCAG	CTGGAAGCGG	GCTCGCACTC	GCATGCCTCC	TCGGGACCCC	3720
	CGAGAGCTGC	CTTGGGCAAG	AGCAGCTATG	GTGTGGCTGC	CCCCGTGGAC	TTCCTGCGCA	3780
	AACAGAGCCA	GCTGCTTCGC	TCGAGGGGCC	CCAGCCAGGC	GGAGCGTGAG	GGCCCGGGAA	3840
5	CACCACCCAC	CACCCTGGCA	CGGGGCAAAG	CACGGAGCAT	CAGCCTGAAG	CTGGACAGCG	3900
	AGGAGTGAGG	CCCACACCAG	CCTGGACCTG	GGTTATTTAT	TGACACACCC	AAGGGGCCCG	3960
	AGGGGCTGCG	TGTGGGGAGG	CTGGGGACCC	AGACTTTTGG	CCCCAGCGCT	GGCCCCCCCCA	4020
	GCCCCACACC	CTATATCTCC	GTGTGCTCCT	CGGTGTTACT	TCCCTTTCAT	ATGAGGGGAC	4080
	CCAGCGCCGG	GGGGAGGGAG	GAGGGCGTGG	GCATGGGCGC	AGAGGCTTTT	CCAGTGTGTA	4140
10	TAAATCCATG	AAAATAAACG	CCACCTGCAC	CCTAAAAAAA	AAAAGTCGAC		4190

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

	(A) LENGTH:	5020 base pairs
	(B) TYPE:	nucleic acid
15	(C) STRANDEDNESS:	single
	(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

	CGGCCGCGT	CGACAAGGAA	CCTTGCCCTAG	AAGTCCCAAC	TTGCAGTTCC	CCATCGACGG	60
	GAAGGCTTGG	ACTCCAAGAT	GATTATAAAG	GAATATCGGA	TTCCTCTGCC	AATGACCGTG	120
	GAGGAGTACC	GCATCGCCCC	GCTGTACATG	ATACAGAAGA	AGAGCCGTAA	CGAGACATAT	180
5	GGCGAAGGCA	GCGGCGTGGA	GATCCTGGAG	AACCGGCCGT	ACACAGATGG	CCGAGGCGGC	240
	TCTGGGCAGT	ACACACACAA	GGTGTATCAT	GTGGGCATGC	ACATTCCCAG	CTGGTTCCGC	300
	TCCATCCTGC	CCAAGGCAGC	CCTGCGGGTG	GTGGAGGAGT	CTTGGAATGC	CTACCCCTAC	360
	ACCCGAACCA	GGTTCACCTG	TCTTTTCGTG	GAGAAATFCT	CCATCGACAT	TGAAACCTTT	420
	TATAAAACTG	ATGCTGGAGA	AAACCCCGAC	GTGTTCAACC	TCTCTCCTGT	GGAAAAGAAC	480
10	CAGCTGACAA	TCGACTTCAT	CGACATTGTC	AAAGACCCTG	TGCCCCACAA	CGAGTATAAG	540
	ACAGAAGAGG	ACCCCAAGCT	GTTCCAGTCA	ACCAAGACCC	AGCGGGGGCC	CCTGTCCGAG	600
	AACTGGATCG	AGGAGTACAA	GAAGCAGGTC	TTCCCCATCA	TGTGCGCATA	CAAGCTCTGC	660
	AAGGTGGAGT	TCCGCTACTG	GGGCATGCAG	TCCAAGATCG	AGAGGTTTAT	CCACGACACC	720
	GGACTACGGA	GGGTGATGGT	GCGGGCTCAC	CGGCAGGCC	GGTGCTGGCA	GGACGAGTGG	780
15	TATGGGCTGA	GCATGGAGAA	CATCCGGGAG	CTGGAGAAGG	AGGCACAGCT	CATGCTTTCC	840
	CGTAAGATGG	CCCAGTTCAA	TGAGGATGGT	GAGGAGGCCA	CTGAGCTCGT	CAAGCACGAA	900
	GCCGTCTCGG	ACCAGACCTC	TGGGGAGCCC	CCGGAGCCCA	GCAGCAGCAA	TGGGGAGCCC	960
	CTAGTGGGGC	GCGGCCTCAA	GAAACAGTGG	TCCACATCCT	CCAAGTCGTC	TCGGTCTGCC	1020
	AAGCGGGGAG	CGAGTCCTTC	CCGCCACAGC	ATCTCAGAGT	GGAGGATGCA	GAGTATTGCC	1080
20	AGGGACTCGG	ATGAGAGCTC	AGATGATGAG	TTCTTCGATG	CGCACGAGGA	CCTGTCCGAC	1140
	ACAGAGGAAA	TGTTCCCCAA	GGACATCACC	AAGTGGAGCT	CCAATGACCT	CATGGACAAG	1200
	ATCGAGAGCC	CAGAGCCGGA	AGACACACAA	GATGGTCTGT	ACCGCCAGGG	TGCCCCTGAG	1260
	TTCAGGGTGG	CCTCCAGTGT	GGAGCAGCTG	AAACATCATAG	AGGACGAGGT	TAGCCAGCCG	1320
	CTGGCTGCAC	CGCCCTCCAA	GATCCACGTG	CTGCTATTGG	TGCTGCACGG	AGGCACCATC	1380
25	CTGGACACAG	GCGCCGGGGA	CCCCAGCTCC	AAGAAGGGCG	ATGCTAACAC	CATCGCCAAAC	1440
	GTGTTTCGACA	CCGTCATGCG	CGTGCACACT	CCACGCGCCC	TGGGCGCGCT	TGCCATCCGC	1500
	CTGGTGCCCT	GCCCCGCCGT	CTGCTCTGAC	GCCTTTGCC	TGGTCTCCAA	CCTCAGCCCC	1560
	TACAGCCATG	ACGAAGGCTG	TCTGTCCAGC	AGTCAGGACC	ACATTCCCCT	GGCTGCCCTC	1620
	CCCCTGCTGG	CCACCTCCTC	CCCCCAGTAC	CAGGAGGCAG	TTGCCACAGT	GATTCAGCGA	1680
30	GCCAACCTTG	CCTATGGGGA	CTTCATCAAG	TCCCAGGAGG	GCATGACCTT	CAATGGGCAG	1740
	GTCTGCCTGA	TTGGGGACTG	CGTCGGGGGC	ATCCTGGCAT	TTGATGCCCT	GTGCTACAGT	1800
	AACCAAGCCG	TGCTTGAGAG	TCAGAGCAGC	AGCCGCGCGG	GCAGCGTGGT	CAGCATGCAG	1860
	GACAAATGACC	TGCTGTCCCC	GGGCATCCTG	ATGAATGCAG	CACACTGTGT	CGGTGGTGGC	1920
	GGTGGCGGCG	GTGGCGGTGG	TGGCAGCAGT	GGTGGTGGTG	GCAGTAGTGG	TGGCTCCAGC	1980
35	CTGGAGAGCA	GTCCGCACCT	GAGCCGAAGC	AACGTCGACA	TCCCCCGCAG	CAACGGCACT	2040
	GAGGACCCCA	AAAGGCAACT	GCCCCGCAAG	AGGAGCGACT	CATCCACCTA	CGAGCTGGAT	2100
	ACCATCCAGC	AGCACCAGGC	CTTCCTGTCC	AGCCTCCATG	CCAGCGTGCT	GAGGACTGAG	2160
	CCCTGTCTAC	GCCATTCCAG	CAGCTCCACC	ATGCTGGATG	GCACAGGTGC	CCTGGGCAGG	2220
	TTTGACTTTG	AGATCACCAG	CCTCTTCCTC	TTCCGGTGCC	CGCTGGGGCT	GGTCTGGCC	2280
40	TTGAGGAAGA	CTGTATCCCC	AGCCCTGGAT	GTTTTCCAGC	TGCGGCCGGC	CTGCCAGCAA	2340
	GTCTACAACC	TCTTCCACCC	CGCGGACCCG	TCAGCTTCAC	GCCTGGAGCC	GCTGCTGGAA	2400
	CGGCGCTTTC	ACGCCCTGCC	GCCTTTTCAGC	GTCCCCCGCT	ACCAACGCTA	CCCGCTGGGG	2460
	GATGGCTGCT	CCACGCTGCT	GCGGGATGTG	CTCCAGACCC	ACAATGCAGC	CTTCCAAGAG	2520
	CATGGCGCCC	CCTCCTCGCC	GGGCACCTGCC	CCTGCCAGTC	GTGGCTTCCG	CCGAGCCAGT	2580
45	GAGATCAGCA	TCGCCAGCCA	GGTGTCAGGC	ATGGCTGAGA	GCTACACGGC	ATCCAGCATC	2640
	GCCCCAGAAG	CCCCCGATGC	GCTCAGCCAT	ACCCCCAGCG	TCAGGCGTCT	GTCCCTGCTC	2700
	GCCCTGCCCG	CCCCCAGCCC	CACCAACCCCT	GGCCCCCACC	CTCCAGCCAG	GAAGGCAAGC	2760
	CCTGGCTGCG	AGAGGGCCCC	TGGCCTCCCT	GAGCTGGACA	TTGGAGAAGT	CGCTGCAAAG	2820
	TGGTGGGGCG	AGAAGCGGAT	CGACTACGCC	CTGACTGCC	CTGACGCCCT	CACGGCCTTC	2880
50	CCCAACGGTG	CTCTGCCCTCA	CCTCTTCCAC	GCCAGCTACT	GGGAGTCAAC	AGACGTGGTC	2940
	TCCTTTCTGC	TGAGACAGGT	CATGAGGCAT	GACAACCTCA	GCATCTTGGA	GCTGGATGGC	3000
	AAGGAAGTGT	CGGTGTTTAC	CCCCCTCAAAG	CCAAGGGAGA	AGTGGCAGCG	CAAGCGGACC	3060
	CACGTGAAGC	TGCGGAACGT	GACGGCCAAC	CACCGGATCA	ATGATGCCCT	TGCCAATGAG	3120
	GACGGCCCCC	AGGTTCTGAC	GGGCAGGTTT	ATGTATGGGC	CCCTGGACAT	GGTCACCCTG	3180
55	ACTGGGGAGA	AGGTGGATGT	GCACATCATG	ACCCAGCCCG	CCTCAGGCCA	GTGGCTCTAC	3240
	CTGGATACGC	TGGTGACCAA	CAACAGTGGG	CGTGTCTCCT	ACACCATCCC	TGAGCTGCAC	3300
	CGCCTGGGCG	TGGGTGTCTA	CCCTATCAAG	ATGGTGGTCA	GGGGAGACCA	CACGTTTGCC	3360
	GACAGCTACA	TCACCGTGCT	GCCCAAGGGC	ACAGAGTTCC	TGGTCTTCAG	CATCGACGGT	3420
	TCCTTTGCGG	CTAGCGTGTC	CATCATGGGC	AGCGACCCCA	AGGTGCGGGC	CGGGGCGGTG	3480
60	GACGTGGTGC	GGCACTGGCA	GGACCTGGGC	TACCTCATCA	TCTACGTGAC	GGGCCGGCCC	3540
	GACATGCAGA	AGCAGCGGGT	GGTGGCGTGG	CTGGCCAGC	ACAAC'TTCCC	CCATGGCGTG	3600
	GTGTCTTTCT	GTGACGGCCT	GGTGCATGAC	CCGTGCGGCG	ACAAGGCCAA	CTTCTCTGAG	3660
	CTGCTCATCT	CCGAGCTGCA	CCTGCGCGTG	CACGCGGCC	ATGGCTCCAC	CAAGGACGTG	3720
	GCGGTGTACA	GCGCCATTAG	CCTGTCCCCC	ATGCAGATCT	ACATCGTGGG	CCGGCCACC	3780
65	AAGAAGCTGC	AGCAGCAGTG	CCAGTTTCATC	ACGGATGGCT	ACGCGGCCCA	CTTGGCGCAG	3840
	CTGAAGTACA	GCCACCGGGC	GCGGCCCGCT	CGCAACACGG	CCACCCGCAT	GGCGCTCGCG	3900
	AAGGGCAGCT	TCGGCCTGCC	CGGCCAGGGC	GACTTTCTGC	GCTCCCGGAA	CCACCTGCTT	3960
	CGACCATCT	CGGCCAGGCC	CAGCGGGCCC	AGCCACCGGC	ACGAGCGGAC	ACAGAGCCAG	4020
	GCGGATGGCG	AGCAGCGGGG	CCAGCGCAGC	ATGAGTGTGG	CGGCCGGCTG	CTGGGGCCCG	4080
70	GCCATGACTG	GCCGCTTGGG	GCCGGGGGCA	GCCGCGGGCC	CCAAGTAGGG	CACCGTGAGT	4140
	GCAGCGCGGG	GTCTCCATGG	TGCTAGGCCA	GGGTGGCCAG	CCCCGCCAGG	AGGCCTGGCC	4200
	TGGGCACACG	CACTGACGTG	GGCCTGGGAG	ATTGTCCCTG	GGCCTTGTGG	AGGACACGGG	4260
	CCGCAACACA	CAGTGCTCCC	TGCCCTGCCT	CACGTCTCCG	GGCCTGACGG	GTCCGGCTTG	4320
	TCATGGAAGC	TGGCAGGGAC	CACCAGCCCC	AGGATGGCAG	AGGGACCAGA	ACCTCCCACT	4380
	CAGACTGGCC	CGGGAGGTTT	TCCCAGACAT	TTTGCCCTGT	GTGGATCTCC	AAGTGTCTCT	4440

	GTGCCAGGTG	TGGGCCCAGG	CGCAGCCTGC	CACCTCCCCA	TCCACTGGCC	ACCCTCACTC	4500
	CCAGGTCCCC	TCCCATTTGG	TAGCAGCTCC	AACAGGGGTC	CAGCCTGCAT	CTTGTTAACT	4560
	CGAGTTTCTC	AACTGTTCAA	CCTCACTGGT	TTTGCACTGA	TTTTTGAGAG	CGGAGACCCA	4620
	TTACCACCTC	CTATGGCTAC	AGCCCCGTTG	ACATGCATGA	AACTCAGTAC	CTGCTGACCC	4680
5	AGGACCTACA	ACCACACTGA	AGGCTCCAGT	GCGGCAGAGC	CTCGTGCAAG	CAGGAGAGAA	4740
	AGGCTGTATC	TTAATTCTG	CACCCCGGAC	CCTGCCCCACC	TGTCTGCCTG	CCCCGCCTGG	4800
	AGCCCAAGCC	AGTGTGTGTT	CCAGCCTCAG	GCCACGGGCT	GGACGGGCCT	GGCCGCCTCT	4860
	TCCGCTCCCT	GCCATCAGTC	AAGGCCGCCC	GCCCACGTTT	CTACGCCTTT	CTACTTCTCA	4920
	ATCTGATTTC	TATGAGGTTT	TTTTAAACGA	GCAATCCTTG	GCTGCTTCCT	TTTCTTAACT	4980
10	CTTTCAGTAC	TGAGAGCAGC	CCCTCCGTCG	ACGCGGCCGC			5020

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

	(A) LENGTH:	1244 amino acids
	(B) TYPE:	amino acid
15	(C) STRANDEDNESS:	single
	(D) TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

MAKAGRAGGPPPGGGAPWHLRNVLSDSVSSDDEFFDAREEMAEGKNAILIGMSQWNS
20 NDLVEQIETMGKLD EHQEGTAPCTSSILQEKQRELYRVSLRRQRFPA
QGSIEIHEDSEEGCPQRSCKTHVLLLVLHGGNILDTGAGDPSCKAADIHTFSSVLEKVTRAHFPAALG
HILIKFVPCPAICSEAFSLVSHLNPYSHDEGLSSSQDHVPLAALPLLAISSPQYQDAVATVIERANQ
VYREFLKSSDGIGFSGQVCLIGDCVGGLLAFDAICY SAGPSGDSPASSSRKGSISSTQDTPVAVEEDC
SLASSKRLSKSNIDISSGLEDEEPPKRPLPRKQSDSSTYDCEAITQHHAFLSSIHSSVLKDESETPAAG
25 GPQLPEVSLGRFDFDVSDFFLFGSPLGLVLAMRRTVLPGLDGFQVRPACSQVYSFFHCADPSASRLEP
LLEPKFHLVPPVSVPRYQRFPLGDGQSLLLADALHTHSPLFLEGSSRDSPLLDAPASPPQASRFQRP
GRRMSEGSSSHSESSSDSMAPVGASRITAKWWSKRIDYALYCPDVLTAFTVALPHLFHASYWEST
DVVAFILRQVMRYESVNIKESARLDPAALSPANPREKWLKRKTQVKLRNVTANHRANDVIAAEDGPQV
LVGRFMYGPLDMVALTGEKVDILVMAEPSSGRWVHLDTEITNSSGRITYNVPRPRRLGVGVYPVKMVV
30 RGDQTCAMSYLTVLPRGMECVVFSIDGSFAASVSIMGSDPKVRPGAVDVVRHWQDLGYMILYITGRPD
MQKQRVVSWLSQHNFPQGMIFFS DGLVHDPLRQKAIFLRNLMQECFIKISAAYGSTKDISVYSVLGLP
ASQIFIVGRPTKKYQTQCQFLSEGYYAAHLAVLEASHRSRPKKNNSRMILRKGSFGLHAQPEFLKRNH
LRRTMSVQQPDPPAANPKPERAQSQPESDKDHERPLPALSWARGPPKFESVP

40

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 1244 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

41

	Met	Leu	Ile	Lys	Glu	Tyr	His	Ile	Leu	Leu	Pro	Met	Ser	Leu	Asp	Glu
	1				5					10					15	
	Tyr	Gln	Val	Ala	Gln	Leu	Tyr	Met	Ile	Gln	Lys	Lys	Ser	Arg	Glu	Glu
				20					25					30		
5	Ser	Ser	Gly	Glu	Gly	Ser	Gly	Val	Glu	Ile	Leu	Ala	Asn	Arg	Pro	Tyr
			35					40					45			
	Thr	Asp	Gly	Pro	Gly	Gly	Ser	Gly	Gln	Tyr	Thr	His	Lys	Val	Tyr	His
	50						55					60				
	Val	Gly	Ser	His	Ile	Pro	Gly	Trp	Phe	Arg	Ala	Leu	Leu	Pro	Lys	Ala
10	65					70					75					80
	Ala	Leu	Gln	Val	Glu	Glu	Glu	Ser	Trp	Asn	Ala	Tyr	Pro	Tyr	Thr	Arg
				85					90						95	
	Thr	Arg	Tyr	Thr	Cys	Pro	Phe	Val	Glu	Lys	Phe	Ser	Ile	Glu	Ile	Glu
				100					105					110		
15	Thr	Tyr	Tyr	Leu	Pro	Asp	Gly	Gly	Gln	Gln	Pro	Asn	Val	Phe	Asn	Leu
			115					120					125			
	Ser	Gly	Ala	Glu	Arg	Arg	Gln	Arg	Ile	Leu	Asp	Thr	Ile	Asp	Ile	Val
	130						135					140				
	Arg	Asp	Ala	Val	Ala	Pro	Gly	Glu	Tyr	Lys	Ala	Glu	Glu	Asp	Pro	Arg
20	145					150					155					160
	Leu	Tyr	His	Ser	Val	Lys	Thr	Gly	Arg	Gly	Pro	Leu	Ser	Asp	Asp	Trp
				165						170					175	
	Ala	Arg	Thr	Ala	Ala	Gln	Thr	Gly	Pro	Leu	Met	Cys	Ala	Tyr	Lys	Leu
				180					185					190		
25	Cys	Lys	Val	Glu	Phe	Arg	Tyr	Trp	Gly	Met	Gln	Ala	Lys	Ile	Glu	Gln
		195						200					205			
	Phe	Ile	His	Asp	Val	Gly	Leu	Arg	Arg	Val	Met	Leu	Arg	Ala	His	Arg
		210					215					220				
	Gln	Ala	Trp	Cys	Trp	Gln	Asp	Glu	Trp	Thr	Glu	Leu	Ser	Met	Ala	Asp
30	225					230					235					240
	Ile	Arg	Ala	Leu	Glu	Glu	Glu	Thr	Ala	Arg	Met	Leu	Ala	Gln	Arg	Met
				245						250					255	
	Ala	Lys	Cys	Asn	Thr	Gly	Ser	Glu	Gly	Ser	Glu	Ala	Gln	Pro	Pro	Gly
				260					265					270		
35	Lys	Pro	Ser	Thr	Glu	Ala	Arg	Ser	Ala	Ala	Ser	Asn	Thr	Gly	Thr	Pro
			275					280					285			
	Asp	Gly	Pro	Glu	Ala	Pro	Pro	Gly	Pro	Asp	Ala	Ser	Pro	Asp	Ala	Ser
	290						295					300				
	Phe	Gly	Lys	Gln	Trp	Ser	Ser	Ser	Ser	Arg	Ser	Ser	Tyr	Ser	Ser	Gln
40	305					310					315					320
	His	Gly	Gly	Ala	Val	Ser	Pro	Gln	Ser	Leu	Ser	Glu	Trp	Arg	Met	Gln
				325						330					335	
	Asn	Ile	Ala	Arg	Asp	Ser	Glu	Asn	Ser	Ser	Glu	Glu	Glu	Phe	Phe	Asp
				340					345					350		
45	Ala	His	Glu	Gly	Phe	Ser	Asp	Ser	Glu	Glu	Val	Phe	Pro	Lys	Glu	Met
			355					360					365			
	Thr	Lys	Trp	Asn	Ser	Asn	Asp	Phe	Ile	Asp	Ala	Phe	Ala	Ser	Pro	Val
	370						375					380				
	Glu	Ala	Glu	Gly	Thr	Pro	Glu	Pro	Gly	Ala	Glu	Ala	Ala	Lys	Gly	Ile
50	385					390					395					400
	Glu	Asp	Gly	Ala	Gln	Ala	Pro	Arg	Asp	Ser	Glu	Gly	Leu	Asp	Gly	Ala
				405						410				415		
	Gly	Glu	Leu	Gly	Ala	Glu	Ala	Cys	Ala	Val	His	Ala	Leu	Phe	Leu	Ile
				420					425					430		

42

	Leu	His	Ser	Gly	Asn	Ile	Leu	Asp	Ser	Gly	Pro	Gly	Asp	Ala	Asn	Ser
			435					440					445			
	Lys	Gln	Ala	Asp	Val	Gln	Thr	Leu	Ser	Ser	Ala	Phe	Glu	Ala	Val	Thr
		450					455					460				
5	Arg	Ile	His	Phe	Pro	Glu	Ala	Leu	Gly	His	Val	Ala	Leu	Arg	Leu	Val
	465					470					475				480	
	Pro	Cys	Pro	Pro	Ile	Cys	Ala	Ala	Ala	Tyr	Ala	Leu	Val	Ser	Asn	Leu
					485					490					495	
10	Ser	Pro	Tyr	Ser	His	Asp	Gly	Asp	Ser	Leu	Ser	Arg	Ser	Gln	Asp	His
				500					505					510		
	Ile	Pro	Leu	Ala	Ala	Leu	Pro	Leu	Ala	Thr	Ser	Ser	Ser	Arg	Tyr	
			515					520					525			
	Gln	Gly	Ala	Val	Ala	Thr	Val	Ile	Ala	Arg	Thr	Asn	Gln	Ala	Tyr	Ser
		530					535					540				
15	Ala	Phe	Leu	Arg	Ser	Pro	Glu	Gly	Ala	Gly	Phe	Cys	Gly	Gln	Val	Ala
	545					550					555				560	
	Leu	Ile	Gly	Asp	Gly	Val	Gly	Gly	Ile	Leu	Gly	Phe	Asp	Ala	Leu	Cys
				565						570					575	
20	His	Ser	Ala	Asn	Ala	Gly	Thr	Gly	Ser	Arg	Gly	Ser	Ser	Arg	Arg	Gly
				580					585					590		
	Ser	Met	Asn	Asn	Glu	Leu	Leu	Ser	Pro	Glu	Phe	Gly	Pro	Val	Arg	Asp
			595					600					605			
	Pro	Leu	Ala	Asp	Gly	Val	Glu	Gly	Leu	Gly	Arg	Gly	Ser	Pro	Glu	Pro
		610					615					620				
25	Ser	Ala	Leu	Pro	Pro	Gln	Arg	Ile	Pro	Ser	Asp	Met	Ala	Ser	Pro	Glu
	625					630					635				640	
	Pro	Glu	Gly	Ser	Gln	Asn	Ser	Leu	Gln	Ala	Ala	Pro	Ala	Thr	Thr	Ser
				645						650					655	
30	Ser	Trp	Glu	Pro	Arg	Arg	Ala	Ser	Thr	Ala	Phe	Cys	Pro	Pro	Ala	Ala
			660						665					670		
	Ser	Ser	Glu	Ala	Pro	Asp	Gly	Pro	Ser	Ser	Thr	Ala	Arg	Leu	Asp	Phe
			675					680					685			
	Lys	Val	Ser	Gly	Phe	Phe	Leu	Phe	Gly	Ser	Pro	Leu	Gly	Leu	Val	Leu
		690					695					700				
35	Ala	Leu	Arg	Lys	Thr	Val	Met	Pro	Ala	Leu	Glu	Ala	Ala	Gln	Met	Arg
	705					710					715				720	
	Pro	Ala	Cys	Glu	Gln	Ile	Tyr	Asn	Leu	Phe	His	Ala	Ala	Asp	Pro	Cys
				725						730					735	
40	Ala	Ser	Arg	Leu	Glu	Pro	Leu	Leu	Ala	Pro	Lys	Phe	Gln	Ala	Ile	Ala
				740					745					750		
	Pro	Leu	Thr	Val	Pro	Arg	Tyr	Gln	Lys	Phe	Pro	Leu	Gly	Asp	Gly	Ser
			755					760					765			
	Ser	Leu	Leu	Ala	Asp	Thr	Leu	Gln	Thr	His	Ser	Ser	Leu	Phe	Leu	
		770				775					780					
45	Glu	Glu	Leu	Glu	Met	Leu	Val	Pro	Ser	Thr	Pro	Thr	Ser	Thr	Ser	Gly
	785					790					795				800	
	Ala	Phe	Trp	Lys	Gly	Ser	Glu	Leu	Ala	Thr	Asp	Pro	Pro	Ala	Gln	Pro
				805						810					815	
50	Ala	Ala	Pro	Ser	Thr	Thr	Ser	Glu	Val	Val	Lys	Ile	Leu	Glu	Arg	Trp
				820					825					830		
	Trp	Gly	Thr	Lys	Arg	Ile	Asp	Tyr	Ser	Leu	Tyr	Cys	Pro	Glu	Ala	Leu
				835				840					845			
	Thr	Ala	Phe	Pro	Thr	Val	Thr	Leu	Pro	His	Leu	Phe	His	Ala	Ser	Tyr
		850				855						860				
55	Trp	Glu	Ser	Ala	Asp	Val	Val	Ala	Phe	Ile	Leu	Arg	Gln	Val	Ile	Glu
	865					870					875				880	
	Lys	Glu	Arg	Pro	Gln	Leu	Ala	Glu	Cys	Glu	Pro	Ser	Ile	Tyr	Ser	
				885						890				895		
60	Pro	Ala	Phe	Pro	Arg	Glu	Lys	Trp	Gln	Arg	Lys	Arg	Thr	Gln	Val	Lys
				900					905					910		
	Ile	Arg	Asn	Val	Thr	Ser	Asn	His	Arg	Ala	Ser	Asp	Thr	Val	Val	Cys
			915					920					925			
	Glu	Gly	Pro	Pro	Gln	Val	Leu	Ser	Gly	Arg	Phe	Met	Tyr	Gly	Pro	Leu
		930					935					940				
65	Asp	Val	Val	Thr	Leu	Thr	Gly	Glu	Lys	Val	Asp	Val	Tyr	Ile	Met	Thr
	945					950					955				960	

43

	Gln	Pro	Leu	Ser	Gly	Lys	Trp	Ile	His	Phe	Gly	Thr	Glu	Val	Thr	Asn
					965					970					975	
	Ser	Ser	Gly	Arg	Leu	Thr	Phe	Pro	Val	Pro	Pro	Glu	Arg	Ala	Leu	Gly
				980					985					990		
5	Ile	Gly	Val	Tyr	Pro	Val	Arg	Met	Val	Val	Arg	Gly	Asp	His	Thr	Tyr
			995					1000					1005			
	Ala	Glu	Cys	Cys	Leu	Thr	Val	Val	Ala	Arg	Gly	Thr	Glu	Ala	Val	Val
		1010					1015					1020				
	Phe	Ser	Ile	Asp	Gly	Ser	Phe	Thr	Ala	Ser	Val	Ser	Ile	Met	Gly	Ser
10		025				1030					1035				1040	
	Asp	Pro	Lys	Val	Arg	Ala	Gly	Ala	Val	Asp	Val	Val	Arg	His	Trp	Gln
				1045					1050					1055		
	Asp	Ser	Gly	Tyr	Leu	Ile	Val	Tyr	Val	Thr	Gly	Arg	Pro	Asp	Met	Gln
			1060						1065					1070		
15	Lys	His	Arg	Val	Val	Ala	Trp	Leu	Ser	Gln	His	Asn	Phe	Pro	His	Gly
		1075					1080					1085				
	Val	Val	Ser	Phe	Cys	Asp	Gly	Leu	Thr	His	Asp	Pro	Leu	Arg	Gln	Lys
		1090				1095					1100					
	Ala	Met	Phe	Leu	Gln	Ser	Leu	Val	Gln	Glu	Val	Glu	Leu	Asn	Ile	Val
20		105				1110					1115				1120	
	Ala	Gly	Tyr	Gly	Ser	Pro	Lys	Asp	Val	Ala	Val	Tyr	Ala	Ala	Leu	Gly
				1125					1130					1135		
	Leu	Ser	Pro	Ser	Gln	Thr	Tyr	Ile	Val	Gly	Arg	Ala	Val	Arg	Lys	Leu
			1140						1145					1150		
25	Gln	Ala	Gln	Cys	Gln	Phe	Leu	Ser	Asp	Gly	Tyr	Val	Ala	His	Leu	Gly
		1155					1160					1165				
	Gln	Leu	Glu	Ala	Gly	Ser	His	Ser	His	Ala	Ser	Ser	Gly	Pro	Pro	Arg
		1170				1175					1180					
	Ala	Ala	Leu	Gly	Lys	Ser	Ser	Tyr	Gly	Val	Ala	Ala	Pro	Val	Asp	Phe
30		185				1190				1195					1200	
	Leu	Arg	Lys	Gln	Ser	Gln	Leu	Leu	Arg	Ser	Arg	Gly	Pro	Ser	Gln	Ala
				1205					1210					1215		
	Glu	Arg	Glu	Gly	Pro	Gly	Thr	Pro	Pro	Thr	Thr	Leu	Ala	Arg	Gly	Lys
			1220				1225					1230				
35	Ala	Arg	Ser	Ile	Ser	Leu	Lys	Leu	Asp	Ser	Glu	Glu				
		1235					1240									

44

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

MIIKEYRIPLPMTVEEYRIAQLYMIQKSRNETYGEESGVEILENRPYTDGPGSGQYTHKVYHVGMMHIPSWFERSILPKA
 10 ALRVVEESWNAYPYTRTRFTCPFVEKFSIDIETFYKTDAGENPDVFNLSPEKNQLTIDFIDIVKDPVPHNEYKTEEDPK
 LFQSTKTQRGPLSENWIEEYKKQVFPIMCAYKLCKVEFRYWMQSKIERFIHDTGLRRVMVRAHRQAWCWQDEWYGLSME
 NIRELEKEAQLMLSRKMAQFNEDGEEATELVKHEAVSDQTSGEPEPSSSSNGEPLVGRGLKKQWSTSSKSSRSSKRGASP
 SRHSISEWRMQSIARDSDESSDDEFFDAHEDLSDTEEMFPKIDITKWSSNDLMDKIESPEPEDTQDGLYRQGAPEFRVASS
 VEQLNIIIEDEVSQPLAAPPKIHVLLLVHGGTILDTGAGDPSSKKGDANTIANVFDTVMRVHYPSALGRLAIRLVPCPP
 15 VCSDAFALVSNLSPYSHDEGCLSSSQDHIPLAALPLLATSSPQYQEAVATVIQRANLAYGDFIKSQEGMTFNGQVCLIGD
 CVGGILAFDALCYSNPVSESQSSSRGVSVMQDNDLLSPGILMNAAHCCGGGGGGGGGGSSGGSSSSLESSRH
 LSRSNVDIPRNGTEDPKRQLPRKRSSTYELDTIQHQAFLLSLHASVLRTEPCSRHSSSTMLDGTGALGRDFEIT
 DLFLFGCPLGLVLALRKTVIPALDVFQLRPACQQVYNLFHPADPSASRLEPLLRFFHALPPFSVPRYQRYPLGDGCSTL
 LADVLQTHNAAFQEHGAPSSPGTAPASRGFRASEISIASQVSGMAESYTASSIAQKAPDALSHTPSVRRLSLLALPAPS
 20 PTTGPHPPARKASPLERAPGLPELDIGEVAKKWQKRIIDYALYCPDALTAFTVALPHLFHASYWESTDVVSFLLRQ
 VMRHDNSSILELDGKEVSFTPSKPREKWQRKTHVKLRNVTANHRINDALANEDGPQVLTGRFMYGPLDMVTLTGEKVD
 VHIMTPPSPGEWLYLDTLVTNNSGRVSYTIPESHRLGVGVPIKMVVRGDHFTFADSYITVLPKGTEFVVFSIDGSFAASV
 SIMGSDPKVRAGAVDVVRHWQDLGYLIIVTGRPDMQKQVVAWLAQHNFPHGVVSCDGLVHDPLRHKANFLKLLISEL
 HLRVHAAYGSTKDVAVYSAISLSPMQIYIVGRPTKKLQQQCQFITDGYAAHLAQLKYSHRARPARNTATRMALRKGSFGL
 25 PGQGDFLRSRNLRLTISAQPSGSPSHRHERTQSQADGEQGRSMSVAAGCWGRAMTGRLEPGAAAGPK

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 986 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

35 Met Leu Ile Lys Glu Tyr Arg Ile Leu Leu Pro Met Thr Val Gln Glu
 1 5 10 15
 Tyr Arg Ile Ala Gln Leu Tyr Met Ile Gln Lys Lys Ser Arg Leu Asp
 20 25 30
 Ser His Gly Gln Asp Ser Gly Val Glu Ile Ile Ser Asn Lys Pro Tyr
 35 40 45
 40 Thr Asp Gly Pro Gly Gly Ser Gly Gln Tyr Thr Phe Lys Ile Tyr His
 50 55 60
 Ile Gly Ser Arg Ile Pro Ala Trp Ile Arg Thr Val Leu Pro Thr Asn
 65 70 75 80

45

	Ala	Leu	Glu	Ala	His	Glu	Glu	Ser	Trp	Asn	Ala	Tyr	Pro	Val	Thr	Lys
					85					90					95	
	Thr	Arg	Tyr	Ser	Thr	Pro	Met	Met	Asp	Arg	Phe	Ser	Leu	Glu	Val	Glu
				100					105					110		
5	Thr	Leu	Tyr	Phe	Asp	Asp	His	Gly	Gln	Gln	Glu	Asn	Val	Phe	Asn	Leu
			115					120					125			
	Asn	Glu	Lys	Asp	Lys	Ser	Thr	Arg	Ile	Ile	Asp	Tyr	Met	Asp	Phe	Val
		130					135					140				
10	Lys	Asp	Pro	Ile	Ser	Ser	His	Asp	Tyr	Cys	Ala	Glu	Glu	Asp	Pro	Lys
	145					150					155					160
	Leu	Tyr	Arg	Ser	Glu	Thr	Thr	Asn	Arg	Gly	Pro	Leu	Asn	Asp	Asp	Trp
				165						170					175	
	Val	Ala	Glu	His	Leu	Lys	Lys	Gly	Leu	Pro	Ile	Met	Cys	Ala	Tyr	Lys
				180					185					190		
15	Leu	Cys	Lys	Val	Glu	Phe	Arg	Tyr	Trp	Gly	Met	Gln	Thr	Arg	Ala	Glu
			195					200					205			
	Arg	Trp	Ile	His	Asp	Leu	Ala	Leu	Arg	Asn	Thr	Met	Met	Arg	Ala	His
		210					215					220				
20	Arg	Gln	Ala	Trp	Ala	Trp	Gln	Asp	Glu	Trp	Thr	Gly	Leu	Thr	Met	Asn
	225				230						235					240
	Asp	Ile	Arg	Lys	Leu	Glu	Ala	Glu	Ala	Ala	Leu	His	Leu	Ser	Lys	Val
				245					250						255	
	Met	Ser	Val	Lys	Glu	Asn	Glu	Asp	Gly	His	Gln	Asp	Glu	Asn	Asp	Thr
				260					265					270		
25	Asp	Asp	Asp	Met	Asp	Ala	Gly	Asp	Ala	Val	Ser	Asp	Asp	Leu	Tyr	Phe
			275				280						285			
	Asp	Cys	Thr	Asp	Thr	Ser	Pro	Ile	Pro	Thr	Gln	Lys	Pro	Ser	Ile	Ile
		290					295					300				
30	Arg	Trp	Ser	Ser	Glu	Leu	Glu	Leu	Glu	Ile	Gln	Asp	Asp	Asn	Ser	Pro
	305					310					315					320
	Pro	Leu	Thr	Pro	His	Asn	Gly	Ser	Thr	Glu	Val	Ala	Leu	Leu	Ile	Met
				325						330					335	
	Val	Phe	His	Gly	Asp	Phe	Ser	Pro	Asp	Asn	Pro	Ala	Asp	Ser	Lys	Thr
				340					345					350		
35	Thr	Asp	Thr	Asn	Thr	Phe	Ser	Ser	Thr	Ile	Glu	Thr	Cys	Val	Gln	Arg
			355					360					365			
	His	Tyr	Pro	Gln	Leu	Arg	Asn	Arg	Leu	His	Ile	Val	Asn	Val	Ser	Cys
				370			375					380				
40	Gly	His	Glu	Met	Thr	Gln	Val	Val	Ser	Lys	Leu	Ser	Asn	Ile	Ser	Pro
	385					390					395					400
	Ser	Phe	Gly	Leu	Leu	His	Pro	Ser	Leu	Ser	Leu	Met	Leu	Pro	Ser	Ala
				405						410					415	
	Ser	His	Leu	Tyr	Asn	Glu	Ala	Val	Glu	Gly	Thr	Ile	Arg	Arg	Ala	Asn
				420					425					430		
45	Glu	Thr	Tyr	Asn	Glu	Phe	Ile	Ala	Ser	Gln	Pro	Leu	Phe	Asn	Gly	Glu
			435					440					445			
	Val	Phe	Val	Val	Gly	Asp	Cys	Val	Gly	Gly	Ile	Phe	Leu	Tyr	Glu	Ala
			450				455					460				
50	Met	Thr	Arg	Lys	Cys	Asp	Ser	Met	Thr	Leu	Leu	Lys	Arg	Leu	Ser	Ser
	465					470					475					480
	Asn	Leu	Ser	Ser	Arg	Ile	Ile	Lys	Glu	Asp	Gln	Ser	Pro	His	Gln	Ser
				485						490					495	
	Met	Thr	Asp	Ile	Thr	Ile	Thr	Asp	Thr	Ser	Ser	Ile	Ser	Ser	Cys	Pro
				500					505					510		
55	Gln	Gln	His	Asn	Gln	Ser	Val	Arg	Asp	His	Ser	Ser	Leu	Gln	Asn	Gly
			515					520					525			
	His	Ala	Ser	Arg	Arg	Ser	Ala	Arg	Asn	Tyr	Ser	Ala	Pro	Pro	Ser	Ala
		530					535					540				
60	Ser	Tyr	Val	Gln	Ile	Asp	Gly	Leu	Asp	Ser	Cys	Gln	Leu	Phe	Asn	Leu
	545					550					555					560
	Tyr	Tyr	Pro	Leu	Asp	Pro	Cys	Gly	Ala	Arg	Ile	Glu	Pro	Val	Leu	Asp
				565						570					575	
	Gly	Gln	Leu	Ser	Cys	Val	Pro	Pro	Tyr	Asn	Val	Pro	Lys	Tyr	Pro	Leu
				580					585					590		
65	Gly	Asp	Gly	Lys	Ser	Gln	Lys	Phe	Glu	Ser	Thr	Ile	Asp	Ala	Thr	Gln
			595					600					605			

46

	Met	Trp	Gly	Ser	Lys	Arg	Ile	Asp	Asn	Leu	Leu	Tyr	Cys	Pro	Asn	Ser
	610						615					620				
	Met	Val	Val	Ala	Leu	Pro	Ser	Ser	Ala	Leu	Pro	Asn	Ile	Leu	His	Ala
	625					630					635					640
5	Ser	Tyr	Trp	Glu	Ser	Cys	Asp	Val	Ala	Ser	Phe	Leu	Leu	Arg	Gln	Phe
					645					650					655	
	Val	Arg	Gly	Glu	Glu	Asn	Ser	Val	Leu	Thr	Thr	Leu	Ser	Ser	Ser	Met
				660					665					670		
10	Asn	Asn	Ile	Pro	Leu	Asn	Ile	Asp	Leu	Pro	Thr	Met	His	Trp	Lys	Arg
			675					680					685			
	Lys	Arg	Thr	Arg	Phe	Lys	Ile	Ala	Asn	Leu	Ser	Ala	Asn	His	Arg	Ala
			690				695					700				
	Asn	Asp	Ile	Leu	Val	Thr	Ala	Gly	Met	Asp	Leu	Thr	Val	Ile	Ala	Lys
	705					710					715					720
15	Phe	Cys	Tyr	Gly	Pro	Met	Asp	Leu	Val	Ala	Leu	Ser	Arg	Glu	Pro	Val
					725					730					735	
	Ser	Val	Phe	Val	Tyr	Pro	Gln	Arg	Gly	Asp	Trp	Tyr	Leu	His	Gly	Val
				740					745					750		
20	Phe	Asp	Thr	Asp	Ser	His	Gly	Arg	Leu	Thr	Leu	Gln	Leu	Ala	Lys	Thr
			755					760					765			
	Leu	Pro	Cys	Gly	Ile	His	Ser	Val	Lys	Ile	Val	Val	His	Gly	Asp	Arg
		770					775					780				
	Ser	Tyr	Leu	Asp	Ala	Phe	Val	Ala	Ile	Val	Pro	His	Gly	Thr	Lys	Cys
	785					790					795					800
25	Ala	Val	Phe	Ser	Val	Asp	Gly	Ser	Leu	Thr	Ala	Ser	Val	Ser	Val	Thr
					805					810					815	
	Gly	Lys	Asp	Pro	Arg	Val	Arg	Pro	Gly	Ala	Val	Asp	Val	Val	Arg	Tyr
				820					825					830		
30	Trp	Gln	Glu	Gln	Gly	Tyr	Leu	Ile	Tyr	Leu	Thr	Ala	Arg	Pro	Asp	
			835					840					845			
	Met	Gln	Gln	Arg	Val	Val	Ser	Ala	Trp	Leu	Ala	Gln	His	Asn	Phe	Pro
		850					855					860				
	His	Ala	Leu	Leu	Phe	Phe	Asn	Asn	Ser	Phe	Ser	Thr	Glu	Pro	Leu	Lys
	865					870					875					880
35	Gln	Lys	Ser	Leu	His	Leu	Arg	His	Ile	Val	Asp	Met	Gly	Val	His	Ile
					885					890					895	
	His	Val	Ala	Tyr	Gly	Ser	Gly	Lys	Asp	Val	Asn	Val	Tyr	Thr	Ser	Ala
				900					905					910		
40	Gly	Val	Asp	Pro	Glu	His	Val	Ile	Ser	Val	Ala	Gly	Ser	Arg	Arg	Arg
			915					920					925			
	Asn	Cys	Val	Gln	Ile	Glu	Ser	Tyr	Ser	Ser	His	Leu	Ala	Ala	Leu	Asn
		930					935					940				
	Ser	Gly	Gln	Cys	Thr	Leu	Gly	Lys	Arg	Ile	Glu	Asp	Asp	Gly	Leu	Thr
	945					950					955					960
45	Leu	Gln	Leu	His	Arg	Asn	Val	Gln	Arg	Thr	Pro	Ser	Phe	Thr	Pro	Arg
					965					970					975	
	Gly	Gly	Lys	Phe	Glu	Asn	Glu	Lys	Asp	Arg						
				980					985							

(2) INFORMATION FOR SEQ ID NO: 8:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4308 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

	GCGGCCGCCA	CAAACAAACA	AACACACGGA	CACACATCTG	GACCTGTACA	CCTACGGCCC	60
	CGGAAAATTA	TCCATAGAAC	AACCGCTGAC	TGACCCCGCC	TCGTTTTTTC	CAATTCCATC	120
	ATTCCGACCA	GGTCATAGAC	GACGTGCCGC	CACCCACGCG	CAATCACCCC	CCTCGCCACA	180
	AAAAACGAAA	AAAAAACCG	TCGACGACA	GCCACGTCGC	GCCTTCACAT	CATCCAGCCA	240
60	TGACCAGCGG	CGGCAATCGA	TGATTGCCAT	TCCCTCAGCC	AACGAGAGCC	AATAGAGGCA	300
	GCCGGAAAGG	AGGACGCCG	AATAGTCAGT	CGGTATCGTC	GGAAGAGTGC	GCCATTGCGA	360

	GAACGTCAAT	AGCCGGAGGG	GAGTCCGCCA	TTTCAACGAC	AAGGACCCAA	GTCACGCGGT	420
	GTCAACATGC	TGATCAAGGA	GTACCGCAT	CCGCTGCCCC	TCACCGTCGA	GGAGTACCGC	480
	ATCGCCCAGC	TCTACATGAT	TGCGAAAAAG	AGTCGCGAGG	AGAGCCATGG	CGAGGGCAGT	540
5	GGCGTTGAGA	TAATCATCAA	TGAGCCGTAC	AAGGATGGAC	CCGGCGGTAA	TGGTCAATAC	600
	ACAAAGAAGA	TCTATCACGT	GGGCAATCAT	CTGCCTGGCT	GGATTAAAAAG	TCTCTTGCCG	660
	AAAAGCGCTT	TAACCGTGGA	GGAGGAGGCC	ATGGAATGCT	ATCCGTATAC	CAGGACTCGC	720
	TACACCTGTC	CGTTTGTGGA	GAAATTCTCG	CTGGATATTG	AGACATACTA	TTATCCGGAC	780
	AATGGCTATC	AGGACAATGT	CTTCCAGCTG	TCCGGAAGCG	ATTTGCGTAA	TCGGATCGTA	840
10	GACGTAATTG	ACATTGTCAA	GGATCAGCTG	TGGGGCGGTG	ACTATGTGAA	GGAGGAGGAT	900
	CCCAAGCACT	TTGTGTCGGA	CAAGACGGGC	CTGGGACCCT	TGGCCGAGGA	TTGGCTGGAG	960
	GAGTATTGGC	GCGAAGTGAA	GGGCAAAAAG	CAACCGACAC	CGCGCAACAT	GTCCCTGATG	1020
	ACCGCCTACA	AGATCTGCCG	CGTGGAGTTT	CGCTACTGGG	GCATGCAGAC	AAAGCTGGAG	1080
	AAGTTTCATCC	ACGATGTGGC	GCTGCGCAAG	ATGATGCTGC	GGGCCCATCG	GCAGGCGTGG	1140
	GCATGGCAGG	ACGAGTGGTT	CGGCTTGACC	ATCGAGGATA	TACGCGAGCT	GGAGCGACAG	1200
15	ACGCAACTGG	CCCTGGCCAA	GAAAATGGGC	GGCGGCGAGG	AGTGACGCGA	CGACAGCGTC	1260
	TCGGAGCCGT	ATGTCAGCAC	GGCGGCCACC	GCCGCATCCA	CAACGGGCAG	CGAGCGAAAG	1320
	AAGTCCGCTC	CGGCTGTGCC	GCCTATTGTC	ACCCAGCAGC	CGCCGAGCGC	CGAGGCCAGT	1380
	TCGGATGAGG	AGGCGGAGGA	GGAGGAGGAT	GACGACGAGG	ACGAGAACGA	TGCCATTGGC	1440
	ACGGGCGTGG	ATCTGTCAGC	CAACCAAGGC	GGATCCGCGC	AGCGCTCGCG	CTCCCAAAGC	1500
20	ATTCAAATGG	CCCAGAAGGG	CAAGTTCGGT	TCAAAGGGTG	CCCTTCACTC	GCCGGTGGGA	1560
	TCTGCCCATTA	GCTTCGATCT	CCAGGTGGCT	AACTGGCGTA	TGGAGCGATT	GGAAGTGGAC	1620
	TCCAAATCCA	ATTCGGATGA	GGAATTCCTT	GATTGCCTGG	ACACCAATGA	GACGAATCTG	1680
	CTGGCCAAGT	GGAGCTCGCT	GGAGCTGCTT	GGCGAGGGCG	ACGACAGTCC	GCCGCCACAT	1740
	GGCGGACCCT	CTAGTGCAGC	ATCGGTGGGT	GGCGTGGCA	ACTCGGGCA	AGAGGACAGC	1800
25	ATATTCAATC	AGGACTTTCT	GATGCGCGTG	GCCCTGGAGC	GCGGCAACAA	CGCGCAGTTA	1860
	CGTTTCTCGG	CCAGCGTGGA	TCGCAGTCAC	GATTTCATCGC	CGCCGGGATC	GCCGAGTACA	1920
	CCGTCTGTGC	CCACAACCAT	TCTGATCCTG	GTGTGCCATG	CGGGCAGCGT	TTTGGATGCG	1980
	GCCAGCGAGC	TGACCGCCAA	GAAATCCGAT	GTGACCAAT	TCCGTGGCTC	CTTCGAGGGC	2040
	GTTATGCGAC	ACGACTATCC	CAGCCTCCTC	ACCCATGTGA	CCATCAAGAT	GGTGCCGTGC	2100
30	CCCTCAATAT	GCACCGACGC	CCTGGGCATT	CTCTCCAGCC	TGAGTCCGTA	CTCCTTTGAT	2160
	GCGTCGCCCT	CGGCGGCGGA	TATACCGAAT	ATAGCCGATG	TCCCCATTGG	AGCTATACCA	2220
	TCTACTATCTG	TGGCATCGCC	AGAATTCAC	GAGACGGTCA	ACAAGACGGT	TGCCGCTGCC	2280
	AATATTGTCT	GCCATGAGTT	TTTGAATTCG	GAGGAGGGTC	ACGGATTCTC	TGGCCAGATT	2340
	GTCAATGCTGG	GCGATTTCAT	GGGTTTCGCTG	CTGGCGTACG	AGGCCCTCTG	CCGATCGAAT	2400
35	GGCAGCCAGC	CGGCGACGGC	TTCGGGTGCC	TCGAATTCGG	GCGGAGATGC	GGCCACAAAT	2460
	ATAAATACCC	ACAATCCGTT	GAGCCCACGT	AATTCGCGAT	TGGACGATGA	CGAGCGTTTC	2520
	ATCGAAGCCG	ATCTGGATGC	CAAGCGTTTG	CTAGTGGCCC	CATCGCCACG	TAGACGCCGT	2580
	TCCAGCTCAT	CCAGCGATTG	GCGTGGCCAC	AAATTGGACT	TTGAGGTCTG	TGACTTCTTC	2640
	ATGTTTCGGAT	CGCCGCTATC	TGTGGTGTCTG	GCTGCAAGGA	AACTTCACGA	TGCCAAGGCC	2700
40	GCCCTGCCCC	GGCCCAACTG	CCACCAGGTC	TACAATCTGT	TCCATCCAAC	CGATCCGATC	2760
	GCCTCGCGCC	TGGAGCCGCT	TCTGAGCGCC	CGGTTTCTA	TATTGGCGCC	AGTCAATGTC	2820
	CCACGGTACG	CCAAGTATCC	GCTGGGTAAT	GGACAGCCAT	TGCATTTATT	GGAGGTCATT	2880
	CAATCGCATC	CGCAGCGCTT	TAACGATGGC	AATAACCTAT	TGGCTGGTGC	CCGTTTGTCTG	2940
	GACGCATCCA	TGCAGAGCAC	GATATCGGGT	CTAGTTGAGA	ATGTCTCGCT	TAGTACGATC	3000
45	CATGCCCTGC	AAAACAAATG	GTGGGGCACA	AAGCGCTTGG	ATTACGCATT	ATATTGCCCC	3060
	GAGGGATTGA	GTAATTTCCC	TGCTCACGCC	TTGCCGCACC	TCTTCCATGC	CAGCTACTGG	3120
	GAGAGTCCGG	ATGTGATTGC	CTTTATTCTA	CGGCAGATTG	GCAAATTCGA	GGGCATACCC	3180
	TTTGTGGGCT	CAAACGATGA	CAAGGACAAT	GCCTCCTTCC	ATCCCGGACA	GCCGAGGGAG	3240
	AAGTGGATTA	AGAAACGGAC	CTCGGTTAAG	CTGAAAAATG	TAGCCGCCAA	TCATCGGGCC	3300
50	AACGATGTAA	TCGTGCAGGA	GGGCAGGGAG	CAGCGATTGA	ATGCGAGATT	TATGTACGGA	3360
	CCCCTGGACA	TGATCACGCT	GCACGGTGAA	AAGGTGGATG	TGCACATTAT	GAAGGATCCG	3420
	CCGGCGGGGC	AGTGGACATT	CCTCAGCACC	GAGGTGACGG	ACAAGAATGG	TCGCATCTCG	3480
	TACAGCATTC	CGGATCAGGT	ATCCCTTGGC	TATGGTATAT	ATCCGGTTAA	GATGGTGGTC	3540
	CGTGGCGATC	ACACCTCGGT	GGATTGCTAT	ATGGCGGTGG	TGCCGCGTTA	ACCGAATGCG	3600
55	TGGTCTTCAG	CATTGATGGC	TCATTACCCG	CTTCGATGTC	GGTGACAGGT	AGGGATCCCA	3660
	AGGTGCGTGC	CGGAGCTGTC	GATGTTTGCC	GCCACTGGCA	GGAGCTGGGC	TACCTGCTCA	3720
	TTTACATCAC	CGGACGACCG	GATATGCAGC	AGCAACGCGT	GGTGCTCTGG	CTGAGCCAGC	3780
	ACAACCTCCC	GCACGGCCTG	ATCTCGTTCC	CCGACGGCCT	GTCCACCGAT	CCATTGGGCC	3840
	ACAAGACGGC	CTATCTCAAC	AATTTGGTTC	AGAACCATGG	AATCTCAATT	ACTGCCCGTA	3900
60	CGGCAGCAGC	AAGGACATTA	GTGTCTACAC	GAATGTTGGC	ATGCGAACCG	ATCAAATTTT	3960
	CATCGTGGGC	AAGGTTGGCA	AGAAGCTGCA	TGCGAATGCC	ACCGTGCTTA	GCGATGGCTA	4020
	TGCGCCCGCT	TTGCGCGGTT	TGCAGGCTGT	GGTGCGGTCG	CGTCCGGCGA	AGGGCAATGC	4080
	CCGCATGGTC	ATTCCACGCG	GATGCTTCAA	TCTTCCCGGC	CAGACCGCAA	ATCCGCGGCG	4140
	CAGAAGGCTG	CATGAACAAG	CAACGAATGA	AAATTGAATT	GCAACTCAAG	CAAACCAATT	4200
65	GTTTAGAGCA	ATGAAAAACA	ACAATTAAAG	CGCTTGTAAG	CAGATAGAAG	ACGTTAAAAAC	4260
	CAAAAAACAA	ACATTACAGA	CAATTGATGT	TAGAATTAGT	GTTCTAGA		4308

48

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 1250 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

10	Met	Leu	Ile	Lys	Glu	Tyr	Arg	Ile	Pro	Leu	Pro	Leu	Thr	Val	Glu	Glu
	1				5					10					15	
	Tyr	Arg	Ile	Ala	Gln	Leu	Tyr	Met	Ile	Ala	Lys	Lys	Ser	Arg	Glu	Glu
				20					25					30		
	Ser	His	Gly	Glu	Gly	Ser	Gly	Val	Glu	Ile	Ile	Ile	Asn	Glu	Pro	Tyr
			35					40					45			
15	Lys	Asp	Gly	Pro	Gly	Gly	Asn	Gly	Gln	Tyr	Thr	Lys	Lys	Ile	Tyr	His
		50					55					60				
	Val	Gly	Asn	His	Leu	Pro	Gly	Trp	Ile	Lys	Ser	Leu	Leu	Pro	Lys	Ser
	65				70						75				80	
	Ala	Leu	Thr	Val	Glu	Glu	Ala	Met	Glu	Cys	Tyr	Pro	Tyr	Thr	Arg	
20				85					90					95		
	Thr	Arg	Tyr	Thr	Cys	Pro	Phe	Val	Glu	Lys	Phe	Ser	Leu	Asp	Ile	Glu
				100					105					110		
	Thr	Tyr	Tyr	Tyr	Pro	Asp	Asn	Gly	Tyr	Gln	Asp	Asn	Val	Phe	Gln	Leu
			115					120					125			
25	Ser	Gly	Ser	Asp	Leu	Arg	Asn	Arg	Ile	Val	Asp	Val	Ile	Asp	Ile	Val
		130					135					140				
	Lys	Asp	Gln	Leu	Trp	Gly	Gly	Asp	Tyr	Val	Lys	Glu	Glu	Asp	Pro	Lys
	145				150					155					160	
	His	Phe	Val	Ser	Asp	Lys	Thr	Gly	Arg	Gly	Pro	Leu	Ala	Glu	Asp	Trp
30				165					170					175		
	Leu	Glu	Glu	Tyr	Trp	Arg	Glu	Val	Lys	Gly	Lys	Lys	Gln	Pro	Thr	Pro
			180						185					190		
	Arg	Asn	Met	Ser	Leu	Met	Thr	Ala	Tyr	Lys	Ile	Cys	Arg	Val	Glu	Phe
		195						200					205			
35	Arg	Tyr	Trp	Gly	Met	Gln	Thr	Lys	Leu	Glu	Lys	Phe	Ile	His	Asp	Val
		210					215					220				
	Ala	Leu	Arg	Lys	Met	Met	Leu	Arg	Ala	His	Arg	Gln	Ala	Trp	Ala	Trp
	225				230						235				240	
	Gln	Asp	Glu	Trp	Phe	Gly	Leu	Thr	Ile	Glu	Asp	Ile	Arg	Glu	Leu	Glu
40				245					250					255		
	Arg	Gln	Thr	Gln	Leu	Ala	Leu	Ala	Lys	Lys	Met	Gly	Gly	Gly	Glu	Glu
			260						265					270		
	Cys	Ser	Asp	Asp	Ser	Val	Ser	Glu	Pro	Tyr	Val	Ser	Thr	Ala	Ala	Thr
			275					280					285			
45	Ala	Ala	Ser	Thr	Thr	Gly	Ser	Glu	Arg	Lys	Lys	Ser	Ala	Pro	Ala	Val
		290					295					300				
	Pro	Pro	Ile	Val	Thr	Gln	Gln	Pro	Pro	Ser	Ala	Glu	Ala	Ser	Ser	Asp
	305				310						315				320	
	Glu	Glu	Gly	Glu	Glu	Glu	Glu	Asp	Asp	Asp	Glu	Asp	Glu	Asn	Asp	Ala
50				325							330			335		
	Ile	Gly	Thr	Gly	Val	Asp	Leu	Ser	Ala	Asn	Gln	Gly	Gly	Ser	Ala	Gln
			340						345					350		
	Arg	Ser	Arg	Ser	Gln	Ser	Ile	Gln	Met	Ala	Gln	Lys	Gly	Lys	Phe	Gly
			355					360					365			
55	Ser	Lys	Gly	Ala	Leu	His	Ser	Pro	Val	Gly	Ser	Ala	His	Ser	Phe	Asp
		370					375					380				
	Leu	Gln	Val	Ala	Asn	Trp	Arg	Met	Glu	Arg	Leu	Glu	Val	Asp	Ser	Lys
	385				390						395				400	
	Ser	Asn	Ser	Asp	Glu	Phe	Phe	Asp	Cys	Leu	Asp	Thr	Asn	Glu	Thr	
60				405					410					415		

49

	Asn	Ser	Leu	Ala	Lys	Trp	Ser	Ser	Leu	Glu	Leu	Leu	Gly	Glu	Gly	Asp
			420						425					430		
	Asp	Ser	Pro	Pro	Pro	His	Gly	Gly	Pro	Ser	Ser	Ala	Ala	Ser	Val	Gly
			435					440					445			
5	Gly	Arg	Gly	Asn	Ser	Arg	Gln	Glu	Asp	Ser	Ile	Phe	Asn	Gln	Asp	Phe
			450				455					460				
	Leu	Met	Arg	Val	Ala	Ser	Glu	Arg	Gly	Asn	Lys	Arg	Gln	Leu	Arg	Ser
						470					475					480
10	Ser	Ala	Ser	Val	Asp	Arg	Ser	His	Asp	Ser	Ser	Pro	Pro	Gly	Ser	Pro
					485					490					495	
	Ser	Thr	Pro	Ser	Cys	Pro	Thr	Thr	Ile	Leu	Ile	Leu	Val	Val	His	Ala
				500					505					510		
	Gly	Ser	Val	Leu	Asp	Ala	Ala	Ser	Glu	Leu	Thr	Ala	Lys	Lys	Ser	Asp
			515					520					525			
15	Val	Thr	Thr	Phe	Arg	Gly	Ser	Phe	Glu	Ala	Val	Met	Arg	His	Asp	Tyr
			530				535					540				
	Pro	Ser	Leu	Leu	Thr	His	Val	Thr	Ile	Lys	Met	Val	Pro	Cys	Pro	Ser
						550					555					560
20	Ile	Cys	Thr	Asp	Ala	Leu	Gly	Ile	Leu	Ser	Ser	Leu	Ser	Pro	Tyr	Ser
					565					570					575	
	Phe	Asp	Ala	Ser	Pro	Ser	Ala	Ala	Asp	Ile	Pro	Asn	Ile	Ala	Asp	Val
				580					585					590		
	Pro	Ile	Gly	Ala	Ile	Pro	Leu	Leu	Ser	Val	Ala	Ser	Pro	Glu	Phe	His
			595					600					605			
25	Glu	Thr	Val	Asn	Lys	Thr	Val	Ala	Ala	Ala	Asn	Ile	Val	Cys	His	Glu
			610				615					620				
	Phe	Leu	Lys	Ser	Glu	Glu	Gly	His	Gly	Phe	Ser	Gly	Gln	Ile	Val	Met
					630						635					640
30	Leu	Gly	Asp	Ser	Met	Gly	Ser	Leu	Leu	Ala	Tyr	Glu	Ala	Leu	Cys	Arg
					645					650					655	
	Ser	Asn	Gly	Ser	Gln	Pro	Gly	Thr	Ala	Ser	Gly	Ala	Ser	Asn	Ser	Gly
				660					665					670		
	Gly	Asp	Ala	Ala	Thr	Asn	Ile	Asn	Thr	His	Asn	Pro	Leu	Ser	Pro	Arg
			675					680					685			
35	Asn	Ser	Arg	Leu	Asp	Asp	Asp	Glu	Arg	Phe	Ile	Glu	Ala	Asp	Leu	Asp
			690				695					700				
	Ala	Lys	Arg	Leu	Leu	Val	Ala	Pro	Ser	Pro	Arg	Arg	Arg	Arg	Ser	Ser
						710					715					720
40	Ser	Ser	Ser	Asp	Ser	Arg	Ala	Thr	Lys	Leu	Asp	Phe	Glu	Val	Cys	Asp
					725					730					735	
	Phe	Phe	Met	Phe	Gly	Ser	Pro	Leu	Ser	Val	Val	Leu	Ala	Ala	Arg	Lys
				740					745					750		
	Leu	His	Asp	Ala	Lys	Ala	Ala	Leu	Pro	Arg	Pro	Asn	Cys	His	Gln	Val
			755					760					765			
45	Tyr	Asn	Leu	Phe	His	Pro	Thr	Asp	Pro	Ile	Ala	Ser	Arg	Leu	Glu	Pro
			770				775					780				
	Leu	Leu	Ser	Ala	Arg	Phe	Ser	Ile	Leu	Ala	Pro	Val	Asn	Val	Pro	Arg
					790						795					800
50	Tyr	Ala	Lys	Tyr	Pro	Leu	Gly	Asn	Gly	Gln	Pro	Leu	His	Leu	Leu	Glu
					805					810					815	
	Val	Ile	Gln	Ser	His	Pro	Gln	Arg	Phe	Asn	Asp	Gly	Asn	Asn	Leu	Leu
					820				825					830		
	Ala	Gly	Arg	Arg	Leu	Ser	Asp	Ala	Ser	Met	Gln	Ser	Thr	Ile	Ser	Gly
			835					840					845			
55	Leu	Ile	Glu	Asn	Val	Ser	Leu	Ser	Thr	Ile	His	Ala	Leu	Gln	Asn	Lys
			850				855					860				
	Trp	Trp	Gly	Thr	Lys	Arg	Leu	Asp	Tyr	Ala	Leu	Tyr	Cys	Pro	Glu	Gly
					870						875					880
60	Leu	Ser	Asn	Phe	Pro	Ala	His	Ala	Leu	Pro	His	Leu	Phe	His	Ala	Ser
					885					890					895	
	Tyr	Trp	Glu	Ser	Pro	Asp	Val	Ile	Ala	Phe	Ile	Leu	Arg	Gln	Ile	Gly
				900					905					910		
	Lys	Phe	Glu	Gly	Ile	Pro	Phe	Val	Gly	Ser	Asn	Asp	Asp	Lys	Asp	Asn
				915				920					925			
65	Ala	Ser	Phe	His	Pro	Gly	Gln	Pro	Arg	Glu	Lys	Trp	Ile	Lys	Lys	Arg
							935					940				

50

Thr Ser Val Lys Leu Lys Asn Val Ala Ala Asn His Arg Ala Asn Asp
 945 950 955 960
 Val Ile Val Gln Glu Gly Arg Glu Gln Arg Leu Asn Ala Arg Phe Met
 965 970 975
 5 Tyr Gly Pro Leu Asp Met Ile Thr Leu His Gly Glu Lys Val Asp Val
 980 985 990
 His Ile Met Lys Asp Pro Pro Ala Gly Gln Trp Thr Phe Leu Ser Thr
 995 1000 1005
 10 Glu Val Thr Asp Lys Asn Gly Arg Ile Ser Tyr Ser Ile Pro Asp Gln
 1010 1015 1020
 Val Ser Leu Gly Tyr Gly Ile Tyr Pro Val Lys Met Val Val Arg Gly
 025 1030 1035 1040
 Asp His Thr Ser Val Asp Cys Tyr Met Ala Val Val Pro Arg Xaa Thr
 1045 1050 1055
 15 Glu Cys Val Val Phe Ser Ile Asp Gly Ser Phe Thr Ala Ser Met Ser
 1060 1065 1070
 Val Thr Gly Arg Asp Pro Lys Val Arg Ala Gly Ala Val Asp Val Cys
 1075 1080 1085
 20 Arg His Trp Gln Glu Leu Gly Tyr Leu Leu Ile Tyr Ile Thr Gly Arg
 1090 1095 1100
 Pro Asp Met Gln Gln Gln Arg Val Val Ser Trp Leu Ser Gln His Asn
 105 1110 1115 1120
 Phe Pro His Gly Leu Ile Ser Phe Ala Asp Gly Leu Val His Asp Pro
 1125 1130 1135
 25 Leu Gly His Lys Thr Ala Tyr Leu Gln Gln Leu Val Xaa Glu Pro Trp
 1140 1145 1150
 Asn Leu Asn Tyr Cys Pro Tyr Gly Ser Ser Lys Asp Ile Ser Val Tyr
 1155 1160 1165
 30 Thr Asn Val Gly Met Arg Thr Asp Gln Ile Phe Ile Val Gly Lys Val
 1170 1175 1180
 Gly Lys Lys Leu Gln Ser Asn Ala Thr Val Leu Ser Asp Gly Tyr Ala
 185 1190 1195 1200
 Ala His Leu Ala Gly Leu Gln Ala Val Gly Gly Ser Arg Pro Ala Lys
 1205 1210 1215
 35 Gly Asn Ala Arg Met Val Ile Pro Arg Gly Cys Phe Asn Leu Pro Gly
 1220 1225 1230
 Gln Thr Ala Asn Pro Arg Arg Arg Leu His Glu Gln Ala Thr Asn
 1235 1240 1245
 40 Glu Asn
 1250

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 10 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

50 Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser
 1 5 10

51

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

10 Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

Claims

1. Isolated, purified, or enriched nucleic acid encoding a rdgB polypeptide.
2. The nucleic acid of claim 1, wherein said rdgB
5 polypeptide is a mammalian rdgB polypeptide.
3. The nucleic acid of claim 2, wherein said mammalian rdgB polypeptide is a human rdgB polypeptide.
4. A nucleic acid probe for the detection of nucleic acid encoding a rdgB polypeptide in a sample.
- 10 5. Recombinant nucleic acid encoding a rdgB polypeptide and a vector or a promoter effective to initiate transcription in a host cell.
6. An isolated, purified, recombinant, or enriched rdgB polypeptide.
- 15 7. The rdgB polypeptide of claim 5, wherein said rdgB polypeptide is a mammalian rdgB polypeptide.
8. The rdgB polypeptide of claim 6, wherein said rdgB polypeptide is a human rdgB polypeptide.
9. A purified antibody having specific binding
20 affinity to a rdgB polypeptide.
10. A hybridoma which produces an antibody having specific binding affinity to a rdgB polypeptide.
11. A method of detecting a compound capable of binding to a rdgB polypeptide comprising the steps of
25 incubating said compound with said rdgB polypeptide and

detecting the presence of said compound bound to said rdgB polypeptide.

12. A method of screening potential agents useful for treatment of a disease or condition characterized by an abnormality in a signal transduction pathway, wherein said signal transduction pathway includes an interaction between a rdgB polypeptide and a natural binding partner, comprising the step of assaying said potential agents for those able to promote or disrupt said interaction as an indication of a useful said agent.

13. A method for diagnosis of a disease or condition characterized by an abnormality in a signal transduction pathway, wherein said signal transduction pathway includes an interaction between a rdgB polypeptide and a natural binding partner, comprising the step of detecting the level of said interaction as an indication of said disease or condition.

14. A method for treatment of an organism having a disease or condition characterized by an abnormality in a signal transduction pathway, wherein said signal transduction pathway includes an interaction between a rdgB polypeptide and a natural binding partner comprising the step of promoting or disrupting said interaction.

15. An isolated nucleic acid molecule comprising a nucleotide sequence that;

(a) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO.:4, SEQ ID NO:5, or SEQ ID NO:6;

(b) the complement of the nucleotide sequence of (a) or;

(c) hybridizes under highly stringent conditions to the nucleotide sequence of (a) and encodes a naturally occurring rdgB protein.

16. A nucleic acid molecule comprising a nucleotide sequence that encodes

(a) a rdgB protein having the full length amino acid sequence of sequence set forth in SEQ ID NO:4 except
5 that it lacks one of the following segments of amino acid residues: 1-616, or 616-974;

(b) the complement of the nucleotide sequence of (a);

(c) a rdgB protein having the full length amino
10 acid sequence set forth in SEQ ID NO:5 except that it lacks one of the following segments of amino acid residues: 1-250, 250-900, or 900-1243;

(d) the complement of the nucleotide sequence of (c);

15 (e) a rdgB protein having the full length amino acid sequence set forth in SEQ ID NO:6 except that it lacks one of the following segments of amino acid residues: 1-251, 251-985, or 985-1349; or

(f) the complement of the nucleotide sequence
20 of (e).

17. A nucleic acid molecule comprising a nucleotide sequence that encodes

(a) a polypeptide having an amino acid sequence set forth in SEQ ID NO:4 from amino acid residues 1-616 or
25 616-974;

(b) the complement of the nucleotide sequence of (a);

(c) a polypeptide having an amino acid sequence set forth in SEQ ID NO:5 from amino acid residues 1-250,
30 250-900, or 900-1243;

(d) the complement of the nucleotide sequence of (c);

(e) a polypeptide having an amino acid sequence of SEQ ID NO:6 from amino acid residues 1-251, 251-985, or
35 985-1349; or

(f) the complement of the nucleotide sequence of (e).

18. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide having the
5 full length amino acid sequence set forth in SEQ ID NO:4; SEQ ID NO:5, or SEQ ID NO:6 except that it lacks at least one, but not more than two, of the domains selected from the group consisting of the PIT domain, the central domain, the PYK2 binding domain, the calcium binding
10 domain and the nucleotide binding domain.

19. A recombinant vector containing the nucleotide sequence of any one of claims 14-18.

20. A genetically engineered host cell containing the nucleotide sequence of any one of claims 14-18.

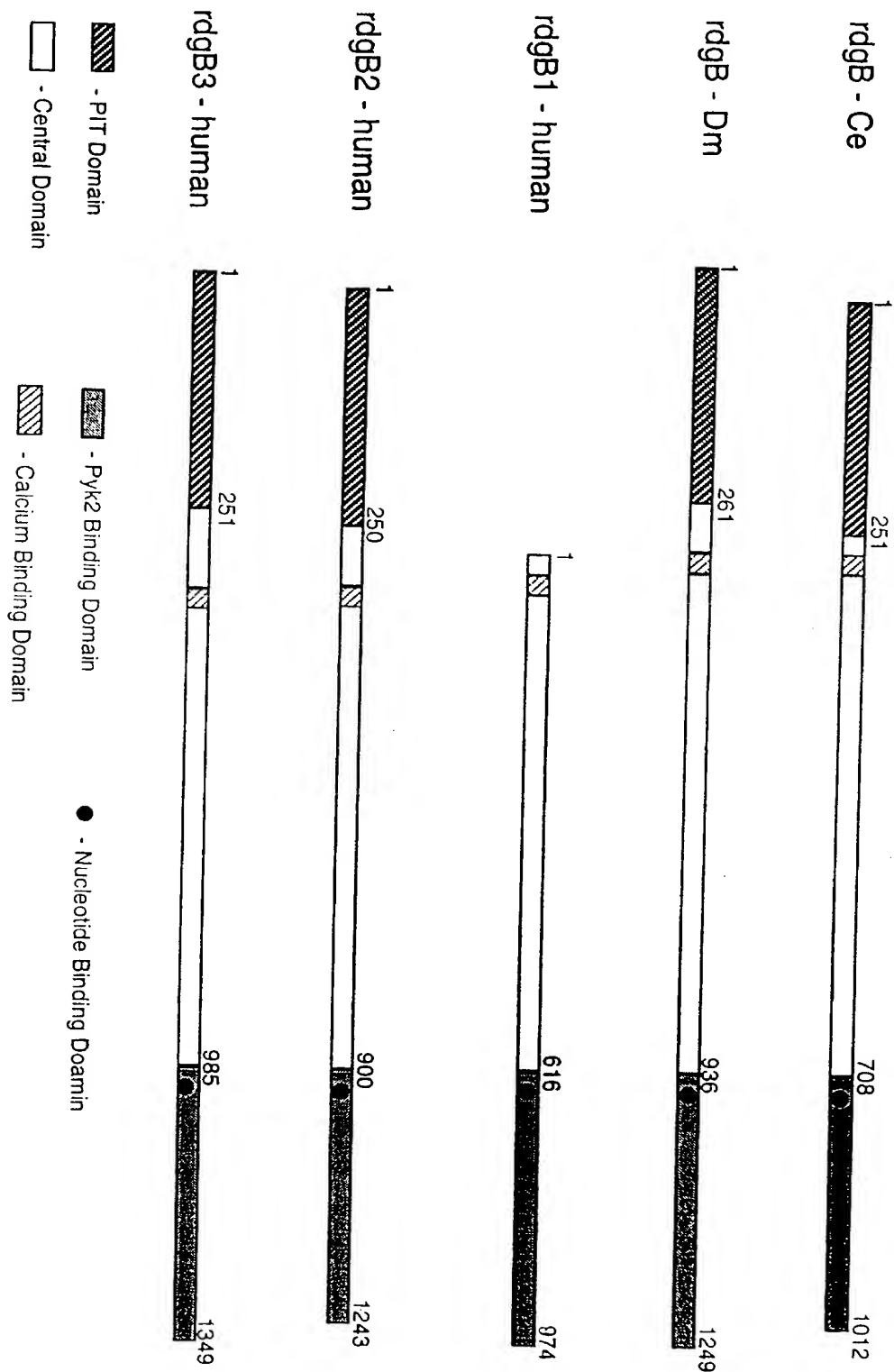


Figure 1

INTERNATIONAL SEARCH REPORT

Internat'l Application No

PCT/US 97/17374

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/435 C07K16/18 C12N5/12 A61K38/17
C12N15/63

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VIHTELIC T S ET AL: "Isolation and characterization of the Drosophila retinal degeneration B (rdgB) gene." GENETICS, vol. 127, April 1991, pages 761-768, XP002055650 --- -/--	1-20



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

13 February 1998

Date of mailing of the international search report

06. 03. 98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Chakravarty, A

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/17374

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	<p> DATABASE EMBL European Bioinformatics Institute Accession Number X98654, 30 June 1997 RUBBOLI F.: "H. Sapiens mRNA for DRES9 protein" XP002055651 see abstract & RUBBOLI F. ET AL.: "A mammalian homolog of the Drosophila retinal degeneration B gene: implications for the evolution of phototransduction mechanisms" GENES FUNCT., vol. 1, 1997, pages 205-214, ----- </p>	1-20